AN ECONOMETRIC ANALYSIS OF FRESH-WINTER VEGETABLE CONSUMPTION: EXTENSIONS OF THE TOBIT MODEL

BY

ANDERSON REYNOLDS

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR IN PHILOSOPHY

UNIVERSITY OF FLORIDA

1989

ACKNOWLEDGEMENT

I wish to express my gratitude to Dr. John S. Shonkwiler, chairman of my doctoral committee, for the guidance and assistance given to me during the preparation of this document. Similarly, I would like to express appreciation to the remaining members of my advisory committee--Drs. Max R. Langham, Jonq-Ying Lee, James A. Zellner and Leonard K. Cheng, for their advice and constructive criticism in the preparation of this document. Special thanks are given to Mercedes Rosalsky for assisting me with data preparation, and to Dr. Robert D. Emerson for serving on my examination committee, on short notice.

Finally, to my parents, I owe the most gratitude. For they instilled in me, at an early age, an appreciation for education, and it is this instilled appreciation for education that has set the stage for past and present academic accomplishments. So, like my master's thesis, I would like to dedicate this document to my parents.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

AN ECONOMETRIC ANALYSIS OF FRESH-WINTER VEGETABLE CONSUMPTION:
EXTENSIONS OF THE TOBIT MODEL

By

Anderson Reynolds

August, 1989

Chairman: Dr. J.S. Shonkwiler
Major Department: Food and Resource Economics

The study utilized cross-sectional data generated from the 1984 Bureau of Labor Statistics Expenditure Diary Survey to analyze the consumption of fresh-winter vegetables. In the process of selecting the censored-regression model most consistent with households' underlying fresh-winter vegetable consumption behavior, three censored-regression models--the Tobit model, Cragg's Double-Hurdle model and the Purchase Infrequency model--were evaluated. Based on the estimated log-likelihoods of the models, the Tobit and Double-Hurdle model appeared to fit the data much better than the Purchase Infrequency model.

Recognizing that misspecification in the form of heteroscedasticity and non-normality yields inconsistent parameter estimates of censored-regression models, the Information Matrix (IM) misspecification test was used to detect violations of the distributional assumptions of the Tobit and Double-Hurdle model. Although methods for parameterizing heteroscedasticity processes exist, accounting for non-normality has been problematic. The study introduces the inverse-hyperbolic-sine transformation as a means for limiting outliers. This transformation has several desirable properties which make its use in the censored-regression context compelling. Likelihood functions which incorporate the transformation are presented. Empirical analysis showed that the Tobit model transformed by the inverse-hyperbolic-sine transformation and with a heteroscedasticity correction yielded a specification that could not be rejected by the IM test at conventional levels of significance. Thus this specification was used to conduct the consumption analysis.

Weekly-household fresh-winter vegetable expenditure was specified as a function of several socioeconomic variables. Among the included variables, food expenditures (income), urbanization, region, season, age, sex, race, education and marital status had a significant impact on fresh-winter vegetable expenditures.

The model was employed to forecast fresh-winter vegetable consumption. Fresh-winter vegetable consumption was projected to increase by 78.9 percent between 1985 and 2010, with population growth and increases in food expenditures accounting for most of the increase.

CHAPTER 1
INTRODUCTION

On a retail-weight basis, fresh vegetables (excluding potatoes) account for over 50 percent of total U.S. vegetable consumption. In 1985 per capita consumption of fresh vegetables amounted to 81.4 pounds, representing 60 percent of the 133.2 pounds of total vegetable consumed (USDA 1986 Agricultural Statistics, Table 686). According to the 1984 consumer expenditure survey, sponsored by the U.S. Department of Labor, out of every dollar American urban consumers spent on food at home approximately 7.8 cents were allocated to vegetables and potatoes. And out of every dollar of such vegetable and potato expenditures, 71 cents were spent on fresh vegetables.

The nutritional contribution of vegetables is well documented. For example, in 1984 vegetables (excluding potatoes) contributed 36.0 percent of U.S. vitamin A intake, 35 percent of the Ascorbic Acid, 11 percent of the Vitamin B_6 and Magnesium, and 9 percent of the Iron intake (USDA 1986 Agricultural Statistics, Table 684).

The vegetable and potato farm enterprises are important sources of farm income. Cash receipts from farm shipments of vegetables (including melons and potatoes) totaled $8.6 billion in 1985, accounting for 11.8 and 6.0 percent, respectively, of crop and total farm shipments (USDA 1986 Agricultural Statistics, Table 583). In 1983, the $53.53 billion marketing cost of fruits and vegetables combined with a farm value of

$13.00 billion resulted in total consumer expenditure on these items of $66.53 billion dollars--21.12 percent of consumer expenditures on farm foods (USDA 1985 Food Consumption, Prices and Expenditures, Table 89).

A significant portion of the fresh vegetables grown and consumed in the U.S. originates in Florida. In fact Florida is second only to California in the production of vegetables. Florida's 1985 production of 1.32 million tons of principal vegetable crops (valued at $570.85 million) represented 5.8 percent of U.S. production. Of the 1.32 million tons, 1.28 million (97 percent) were produced for the fresh market. This allocation to the fresh market accounted for 11.7 percent of all such allocations (USDA 1986 Agricultural Statistics, Tables 199-200).

Table 1 indicates that Florida's fresh vegetable shipments over the crop year are unevenly distributed. For example, during the 1986-1987 crop year, 22.5 percent (7.4 million cwt) of fresh vegetable marketings of the eight major vegetable crops (snap beans, celery, sweet corn, cucumbers, lettuce, green peppers, squash, and tomatoes) was shipped in the month of May. The months of April, December and November followed with shipments of 5.4, 4.2 and 3.9 million cwt, respectively, accounting for 16.3, 12.8 and 11.7 percent of all such shipments. With shipments of 0.6 and 0.1 million cwt, respectively, the months of October and July accounted for the least amount (1.8 and 0.4 percent) of fresh vegetable sales. An examination of previous crop years reveals that this sale distribution pattern observed for the 1986-1987 crop year holds true historically. Average monthly shipments of these vegetable crops over five crop years (1982-1983 through 1986-1987) are also

Table 1. Monthly Sale Distribution of Florida's Principal Fresh Vegetable Crop, 1982-83 to 1986-87.

Crop Year	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Total
						1000 Cwt					
1982-83	558	3,024	4,013	3,525	2,183	2,424	4,245	6,481	4,736	349	31,538
1983-84	505	3,348	4,538	2,248	2,095	3,796	5,220	6,622	2,982	181	31,536
1984-85	692	4,437	4,050	3,845	1,750	1,924	6,142	8,009	1,863	74	32,786
1985-86	562	4,152	3,890	3,337	2,370	3,079	5,758	7,862	2,490	65	33,566
1986-87	640	4,331	4,574	3,518	2,697	3,157	5,462	7,925	2,834	129	35,269
Average	591	3,858	4,213	3,295	2,219	2,876	5,366	7,380	2,981	160	32,939
						Percent					
1982-83	1.8	9.6	12.7	11.2	6.9	7.7	13.5	20.5	15.0	1.1	100
1983-84	1.6	10.6	14.4	7.1	6.6	12.0	16.6	21.0	9.5	0.6	100
1984-85	2.1	13.5	12.4	11.7	5.3	5.9	18.7	24.4	5.7	0.2	100
1985-86	1.7	12.4	11.6	9.9	7.1	9.2	17.2	23.4	7.4	0.2	100
1986-87	1.8	12.3	13.0	10.0	7.6	9.0	15.5	22.5	8.0	0.4	100
Average	1.8	11.7	12.8	10.0	6.7	8.7	16.3	22.4	9.0	0.5	100

Source: Florida Department of Agriculture, Agricultural Statistics Service.

presented in Table 1. According to these averages, the months of May, April, December and November, in order of importance, had the greatest share of fresh vegetable shipments, while July and October had the least amount.

Given Florida's share of U.S. vegetable production, it is not surprising that the vegetable industry plays an important role in the state's agricultural economy. According to the Unemployment Compensation Law, in 1986 there was a monthly average of 395 reporting establishments engaged in the production of vegetables and melons in the state of Florida. These establishments employed more than 24 thousand workers to whom they paid a total of $15.54 million in wages. In 1984, cash receipts from farm marketings of vegetables and potatoes amounted to $1.01 billion, representing over 20 percent of the state's farm income (Florida 1987 Statistical Abstract, Tables 9.14, 9.26).

Problem Statement

This study is concerned with the impact or relative impact of various socioeconomic and demographic factors on U.S. consumption of fresh-winter vegetables. Such information can be used to forecast or project consumer expenditures for planning and decision-making purposes. Policy makers can use knowledge of how income and demographics affect food consumption to assess or anticipate the dietary effects of assistance programs.

An analysis of the impact of demographic variables on the consumption of fresh-winter vegetables normally involves the use of cross-sectional data on individual household characteristics along with the household's fresh-winter vegetable expenditures. However, surveys designed to obtain such data often include a large number of households which did not report any expenditures. Consequently, standard regression methods provide an inappropriate framework for conducting the demand analysis. Recognizing this, several researchers (Capps and Love 1983, Smallwood and Blaylock 1984, Blaylock and Smallwood 1986) have employed the Tobit model to analyze U.S. vegetable consumption.

Under certain conditions, however, the traditional Tobit model may produce inconsistent results. Specifically, if the disturbance term is non-normally distributed or exhibits heteroscedasticity the estimated parameters would be inconsistent (Hurd 1979, Goldberger 1983). Furthermore, the Tobit model assumes that zero expenditures are observed when desired expenditures are non-positive; thus, the dependent variable is truncated at zero. However, as Maddala (1985) has pointed out, if zero expenditures are a reflection of the choice of consumers rather than the unobservability of desired expenditures, the Tobit model would be a misrepresentation of the underlying consumption behavior.

In this study the Tobit model along with other models that provide alternative explanations for the occurrence of zero expenditures were evaluated. In addition, misspecification tests were conducted, and appropriate measures were taken in an effort to improve model specification.

Objectives

In general this study was concerned with an analysis of the impact of socioeconomic and demographic variables on U.S. consumption of fresh-winter vegetables. Specific objectives were to

1. evaluate alternative censored regression models in an attempt to select a model that is consistent with household's fresh-winter vegetable consumption behavior;
2. conduct misspecification testing and model transformations to adjust for apparent misspecifications;
3. forecast U.S. fresh-winter vegetable consumption.

Data

The study utilized data from the Continuing Consumer Expenditure Survey (CCES) sponsored by the Bureau of Labor Statistic (BLS), U.S. Department of Labor. The CCES represents a recent, comprehensive data set on food spending by Americans (Smallwood and Harris, 1987) and consists of two separate parts: (1) a quarterly interview panel survey in which approximately 5000 consumer units (households) are interviewed every three months over a 15-month period, and (2) a diary or recordkeeping survey completed by each consumer unit in the sample for two consecutive one-week periods. Only the diary survey is used in the present study.

The primary focus of the diary survey is to obtain expenditure data on small, frequently purchased items that do not lend themselves to easy recall. Hence, during the two consecutive one-week survey periods each household was asked to record its expenditures on such items as food, beverages, tobacco, housekeeping supplies, nonprescription drugs, personal care products and services, fuels, and utilities. The survey, however, excluded purchases made while away from home overnight, purchases for business use, and credit payments on goods and services already acquired.

In addition to the household expenditure data, at the beginning of the survey period the Census interviewer used a household characteristic questionnaire to record information on the age, sex, race, marital status, education and family relationships of members of the consumer unit. And, at the end of the survey period, the same household characteristics questionnaire was used to collect previous-year information on the work experience, occupation, industry, retirement status, earnings from wages and salaries, net income from one's own farm, income from other sources, and other household characteristic data.

To obtain some insights into the nature of the data regarding how expenditures on fresh vegetables defer across socio-demographic characteristics, average weekly expenditures by various socio-demographic groupings are tabulated in Table 2. According to the data, American households spend an average weekly amount of $1.52 on fresh vegetables. Such expenditures vary directly with household size. For example, while a one-person household spends only $0.73 cents on fresh vegetables, a five-person household spends over $2.00, and households with six or more occupants spend $2.92 on fresh vegetables. Up to a certain age (65 years), expenditures on fresh vegetables also appear to be directly related to the age of the household head. Expenditures increases continuously from a low of $0.60 associated with households whose heads are under 25 years to a high of $1.95 associated with households whose heads are between the ages of 55 and 64 years. According to the data, male headed households spend $1.65 on fresh vegetables compared with an expenditure level of $1.23 for female headed

Table 2. Household Expenditures on Fresh Vegetables by various Household Characteristics, 1984.

Household Characteristics	Number of Households	Average Weekly Expenditures
All Households	3368	1.5132
Household Size		
One person	937	0.7302
Two persons	983	1.5870
Three persons	552	1.7128
Four persons	492	1.9505
Five persons	258	2.0377
Over five persons	144	2.9170
Age of Reference Person		
Under 25	342	0.5975
25 - 34	729	1.2177
35 - 44	619	1.8208
45 - 54	488	1.9320
55 - 64	550	1.9486
over 65	640	1.3481
Sex of Reference Person		
Male	2256	1.6539
Female	1112	1.2278
Race		
White	2879	1.5102
Black	386	1.0638
Nonwhite/nonblack	103	3.2820
Education of Reference Person		
High school graduate	2455	1.5542
Not high school graduate	913	1.4029
Marital Status		
Married	1937	1.9066
Single	1431	0.9807
Location		
Urban	3006	1.5509
Rural	362	1.2006
Region		
Northeast	1052	1.5260
Midwest	795	1.2956
South	798	1.4445
West	723	1.8096

Source: U.S. Department of Labor, Bureau of Labor Statistics.

households. With regard to race, households whose heads are nonwhite/nonblack spend the most ($3.28) on fresh vegetables, followed by white households with an expenditure level of $1.51. In contrast, households headed by blacks spend only $1.06 on fresh vegetables. Expenditures are also dependent on the educational level and marital status of the household head. Households headed by high school graduates spend $1.55 on fresh vegetables, while those headed by non-high-school graduates spend $1.40. Households with married couples spend $1.91 on fresh vegetables, almost $1.00 more than what households without married couples spend. The data indicate that fresh vegetable expenditures differ across location, both in terms of urbanization and region. For example, urban dwellers spend about $1.55 on fresh vegetables while rural households spend only $1.20. With regard to region, households located in the West spend the most ($1.81) on fresh vegetables. Northeastern households follow with an expenditure level of $1.53, while households located in the South and in the Midwest spend $1.44 and $1.30, respectively, on fresh vegetables.

CHAPTER 2
LITERATURE REVIEW

The literature review for this study falls into three main sections: first, a review of past studies concerned with the incorporation of demographic factors into demand functions (or Engel curves); second, studies that have introduced economic and demographic variables in the analysis of vegetable consumption will be examined; and finally, studies that have dealt with the specification and testing of censored-regression models, within the single equation framework, will be explored.

Demographic Effects in Demand Equations

The neoclassical theory of consumer behavior suggests that given a household's preferences for good and services satisfy certain regularity conditions (Varian 1984, pg. 113), there exists a continuous utility function which represents those preferences. The theory then assumes that the household maximizes utility subject to a budget constraint. Given this behavioral assumption and a well behaved utility function, the derivation of the consumption bundle that is consistent with utility maximization behavior is reduced to a mathematical problem. This optimal consumption bundle is usually specified as a function of prices and income. The Engel curve, which relates household expenditures to household income while prices are held constant, is a special case of the solution to the maximization problem discussed above. The notion of an Engel curve has been in vogue since the discovery made by Engel (1895) that the poor allocate a larger share of their income to food than do the rich.

The majority of demand or Engel curve specifications aggregate over consumers or households. However, the neoclassical theory of consumer behavior is based on individual decision units. Recently survey data on individual households have become more readily available and have thus facilitated demand studies.

There have been numerous modifications or extensions of the neoclassical conceptualization of consumer behavior (Ferber 1973). One such extension is based on the realization that there are many factors other than price and income that influence consumer preferences and hence their choice set. One group of variables--household composition and other demographic variables--have attracted a great deal of attention, as determinants of the consumption pattern of households. Demographic factors influencing preferences have intuitive appeal because the needs of households differ along with household characteristics. For example, a household of equal size but with younger children than another household would be expected to need less food. Furthermore, there are economies of scale in consumption. Larger households waste less food and purchase in larger quantities and hence require proportionately fewer food expenditures than a smaller counterpart. This notion can be further extended to other demographic factors such as level of education, race or national origin, and geographical location. These factors in one way or the other (tradition, habit persistence, level of appreciation of the nutritional content of foods in the case of educational level, etc.)

condition one's preferences and hence one's perceived needs. The neoclassical theory and the traditional concept of demand functions or Engel curves neglect the variation in need arising from age and other household characteristics, and also the opportunities for economies of scale in consumption. Realizing this, Engel conceived the notion of household equivalent scales which can be construed as index numbers that correct for such variation in needs. This was accomplished by expressing the needs of each household with reference to a representative household, thus obtaining a specific scale or weight for each household as a function of various household characteristics. The introduction of equivalent scales gave rise to utility functions which have as arguments commodities and household characteristics. From such a utility function, demand functions which specify individual commodities deflated by household specific scales as functions of prices and income deflated by the same household specific equivalent scales can be obtained.

Sydenstricker and King (1921), followed by Prais and Houthakker (1971), recognized that Engel's model wrongly assumed that the needs of children relative to adults and the economies of scale in consumption were the same for every commodity. For example, while a child will most likely need considerably fewer cigarettes than an adult, we can expect that same child to consume nearly as much or even more ice cream than the adult. To take into account this commodity specific effect, these authors generalized the Engel curve by specifying each individual commodity deflated by its own commodity specific scale as a function of household income deflated by a composite scale defined as a weighted average of scales specific to each commodity.

Barten (1964) has provided a different generalization of Engel's model. He specified the household's utility function as a function of commodities deflated by their own equivalent scales. Muellbauer (1974) has shown that this utility function gives rise to individual marshallian demand functions expressed as a product of the commodity's own equivalent scale and a function which has as arguments ratios (one ratio for each good in the consumption bundle) of household income to each commodity price weighted by the commodity's equivalent scale. This generalization, as opposed to the previous, is considered to be a true generalization of Engel's model. Engel's model took into consideration only the absolute effect of household composition on prices--additional children in the household necessarily imply additional expenditures on children related goods, thus an increase in the price of these goods to the household. In contrast, Barten's model, in addition to the absolute price response, included a substitution effect--as the price of children goods increases relative to other goods there is a substitution away from children goods to other goods. The Sydenstricker-King and the Prais-Houthakker model is not considered a true generalization of Engel's model because their model does not include a substitution term. In fact their model is consistent with the theory of consumer behavior (or is identical to the Barten model) only in the case where the utility function permits no substitution between goods (Muellbauer 1974).

Engel's noncommodity specific adult equivalent scales and Barten's commodity specific equivalent scales represent means of introducing demographic variables into demand equations. Barten's method has been named demographic scaling (Pollak and Wales 1980), because this method allows both preferences and demand behavior to be viewed in terms of demographically scaled prices and quantities (scaled or deflated by the commodity specific adult equivalent scale). Demographic translating, which was first introduced by Pollak and Wales (1978), is another procedure for incorporating demographic variables into demand equations. This method modifies demand systems by introducing parameters, which are functions of demographic variables, additively into the original demand system. Gorman (1976) has proposed a more general specification of which demographic scaling and translating are special cases. In addition to including translating parameters, Gorman's method includes commodity specific equivalent scales in much the same way as Barten's model.

More recently Lewbel (1985) has pointed out that the above methods of incorporating demographic effects into demand systems are restrictive because these methods rule out complicated interactions of demographic variables with prices and/or total expenditures (income). As an alternative he suggested modifying functions which constitute an even more general method of introducing demographic or other effects into demand systems. This method involves the introduction of functions (modifying functions) of demographic variables, prices and income into the expenditure function of demand systems. Modifying functions do not only permit scaling and translation terms to be functions of expenditure levels and demographic variables, but also allow considerable interaction between demographic factors and both price and income.

Apart from the above systematic methods of introducing demographic variables into demand analysis, more ad hoc methods exist. One such method is called unpooled estimation (Kokoski 1986). In this method the demand system is estimated separately for each demographic group, thus obtaining a set of parameter estimates for each demographic group. Kokoski indicated that this method assumes that all demand parameters may be affected by demographic factors with no prior specification of the relationship between the parameters and demographic effects. Another method that is commonly used is the inclusion of demographic variables on the right hand side of demand equations for single goods. As Lewbel has pointed out, this method allows for virtually any set of demographic and price effects but does not have general applicability, being specific to the model at hand. However, this method avoids the added complexity of specifying adult equivalent or commodity specific scales.

The above studies have treated demographic variables as exogenous to the utility maximizing process. Chavas and Citzler (1988), within the framework of Barten's model and borrowing from Becker's (1965) household production theory, have introduced the novel approach of endogenizing demographic factors in the analysis of consumer behavior. In brief, they assumed that various socio-demographic factors such as household size, age and educational level are indications of the amount of human capital the household possesses. Therefore, since human capital can be expected to directly influence the production of both market goods (income) and non-market goods, demographic factors can be considered to be indirect inputs in the production of such goods. Also, since the household must decide how many resources should be allocated to the production of income in order to maximize utility, income and hence the demographic variables that influence income are endogenous to the utility maximizing process. Notwithstanding the contribution of household composition to household income, there is a cost associated with household composition (the addition of an additional child for example). Thus, since household composition influences income, within a long-run framework, the household chooses the household composition such that the marginal cost of a family member is equal to its marginal revenue. It is exactly this notion, that household composition is in part determined within the utility maximizing process, that marks the point of departure of Chavas and Citzler's study from previous studies that have attempted to introduce demographic variables into demand analysis.

The Impact of Economic and Demographic Factors on Vegetable Consumption

Several studies have analyzed the impact of various socioeconomic and demographic factors on the consumption of vegetables. Buse and Salathe (1978), with data from the 1955 and 1965 USDA household food consumption surveys, employed adult equivalent scales to incorporate household composition effects into Engel curve specifications for various food groups. They specified the scales as continuous functions of the age of household members, with the restrictions that at age zero the value of the scale is the same for both male and female and there after allowed to be different. In addition, while the scale is allowed to change between the ages of 0 and 20 and between 55 and 75, the value of the scale is constrained to constancy between the ages of 20 and 55 and over the age of 75. Along with adult equivalent scales, the number of meals away from home, and the race, education and employment status of the female head were included as explanatory variables. Also included were household income, the square of the number of adult equivalent and a number of interaction variables, including the number of adult equivalent in the household interacting with region, urbanization, and the race of the female head, and income interacting with the race, education, and employment status of the female head. The food expenditure equations were then estimated with a nonlinear regression algorithm. The results indicated that the marginal propensity to spend on vegetables varies indirectly with the level of education of the female head and her labor market participation. Households located in the South spend the least per adult equivalent while those residing in the Northeast exhibited the greatest tendency to spend on vegetables. In addition, rural households spend less per adult equivalent than their urban counterparts. Other results suggest that the addition of a newborn baby, or an adult female, or an elderly member, had a significant positive impact on vegetable consumption. In addition to estimating the Engel curve functions, statistical tests were performed on the adult equivalent scale parameters. The tests revealed that adult females predisposes the household to spend more on vegetables than both female children and elderly females. Similarly, the presence of adult males predisposes the household to spend more than male children. On the other hand, the sex of household members was found to be an insignificant determinant of vegetable expenditures.

Salathe (1979a) used data from the 1965-66 USDA Food Consumption Survey to analyze the effects of changes in population characteristics on U.S. consumption of selected foods. He isolated the effect of age and sex on food intake by partitioning individual records into twenty different age-sex groups. Next Salathe used regression analysis to isolate the impact of household size, racial mix, regional population shifts and urbanization on food intake. Each selected food item was regressed (one equation for each age-sex group) against 1950 census estimates of these variables. After these 1950 per capita consumption estimates were obtained, the effect of each individual independent variable on consumption in subsequent years was estimated by holding all other variables constant at their 1950 level. These estimates were then used to compute indices (with 1950 as the base year) which were construed as percentage change in per capita consumption in response to changes in the particular explanatory variable. Changes in age-sex composition were estimated to have caused vegetable consumption to increase by 2.9 percent between 1960 and 1975. However, based on projected changes in age-sex composition, vegetable consumption was predicted to decrease by 0.2 percent from 1975 to 1990. Because declines in household size are accompanied with increased per capita income, according to the study, declining household sizes since 1960 have had a positive impact on virtually all food groups. In the case of vegetables, such changes in household size were estimated to bring about a 0.7 percent increase in per capita consumption between 1960 and 1990. Race is another factor considered to influence food consumption. The author indicated that the data used for the study revealed that blacks and other minorities as a group consume smaller quantities of vegetables on a per capita basis than whites. Not surprisingly, since blacks and minorities share of the population is on the increase, the study indicates that changes in racial mix may cause vegetable consumption to decrease by 0.5 percent over the 1960-1990 period. The study revealed that regional shifts in population have no apparent impact on vegetable consumption. In contrast, the rural to urban migration trend is estimated to cause per capita consumption to decrease by 0.8 percent between 1960 and 1990.

Another study by Salathe (1979b) was based on data generated from the 1972-73 Bureau of Labor Statistics Consumer Expenditure Diary survey. Ordinary least squares were applied to food expenditure functions quadratic in household size and household income. An income elasticity for fresh vegetables was estimated at 0.19, while income elasticities for frozen and other processed vegetables (canned and dried vegetables) were estimated to be 0.43 and 0.03, respectively. Except for the frozen vegetable category, household size elasticities for all vegetable categories were consistently larger than their corresponding income elasticities. In fact, household size elasticities for vegetables ranged from a low of 0.40 for frozen vegetables to a high of 0.77 for potatoes. Smallwood and Blaylock (1981) used the same model specification and estimation procedure as Salathe, but with data generated by the USDA 1977-78 Nationwide Food Consumption Survey.

Except in the case of canned vegetables for which they obtained a negative income elasticity of 0.10, the results of Smallwood and Blaylock were very similar to those of Salathe.

In 1980 Price, Price and West used regression analysis to determine the effect of traditional factors such as household size, composition and location along with nontraditional variables such as liquid assets, household management style and psychological need levels, on both the level and variety of fruits and vegetables consumed by Washington households. The data used for the study were collected from the state of Washington during 1972 and 1973, from a sample of 497 white households containing 8- to 12-year-old children. According to the results, liquid assets have a significant positive effect on fresh vegetables but do not seem to influence processed vegetables. In contrast, current income had a significant but negative impact on fresh garden and Mexican vegetables. Reasons given for these unexpected results were that the Mexican vegetable grouping contains inexpensive foods and the fresh vegetables may be reflecting home garden production among low-income households. With the exception of the fresh green vegetable category, household size had a positive and statistically significant impact on all fresh vegetables. Household size also positively influenced processed vegetables but the results were less convincing. The educational level of the adult female was also an important determinant of vegetable consumption. This variable had a significant and positive influence on the common fresh vegetables category, and a negative but almost equally significant impact on the common frozen vegetable category. Although the occupation of the major earner lacked explanatory power, the results indicated that white collar workers tend to consume more green vegetables, both in fresh and frozen forms, than do others.

Regional differences in food consumption patterns among low income households was the main focus of a study by Matsumoto (1984). Using the low-income supplemental survey (of about 4600 households) generated by the Nationwide Food Consumption Survey, he regressed seven food expenditure groups on various socioeconomic and demographic variables for each of nine regions. The results indicated that low income household's food consumption responses to changes in income differ considerably across regions. The West North-Central, South Central and Mountain states, with marginal propensities to consume fruits and vegetables ranging from 2.03 to 2.96, were the most responsive. Next were the East North-Central, South Atlantic and Pacific states with marginal propensities to consume fruits and vegetables of between 1.01 and 1.54. Finally, the New England and Mid-Atlantic states, with marginal propensities of less than 1.0, showed the least tendency to consume fruits and vegetables given a change in income. A comparison of regions with regard to income elasticities for fruits and vegetables exhibited the same pattern as the marginal propensity to consume. The above first group of states had income elasticities of over 0.31, the second group exhibited elasticities ranging from 0.14 to 0.22, and the least responsive group of states had elasticities of less than 0.12. In contrast to the study by Smallwood and Blaylock (1981), family size (household size) had a negative impact on fruit and vegetable expenditures in all regions and for the nation as a whole. However, Matsumoto dealt with only low income households.

Recently, the Tobit model has seen some application in the analysis of vegetable consumption. Huang, Fletcher and Raunikar (1981), with the 1972-73 Consumer Expenditure Diary Survey data, used the Tobit model to analyze the effects of the food stamp program on participating households' food purchases. Among the explanatory variables included in the analysis, household income, the degree of participation, the race of the household head, the region in which the household resided and also the rural/urban location of the household, significantly influenced participating households' fruit and vegetable expenditures. Expenditures on fruits and vegetables were found to be positively related to household income (however, income elasticity was inelastic), urban versus rural residents, full participants of the program as oppose to partial participants, whites as opposed to other races and households residing in the Northeast as opposed to other parts of the country.

The study by Capps and Love (1983) represents a second study which employed the Tobit model in the analysis of vegetable consumption. Data from the 1972-1974 Consumer Expenditure Dairy Survey were used to examine the impact of socioeconomic factors on fresh vegetable expenditures. Apart from income, the study included as explanatory variables, household age sex composition, household earner composition, education of household members, race of household head, household food stamp participation, population density, and the region in which the household is located. The study reported an income elasticity of 0.24

for fresh vegetables. Other results indicate that economies of scale in consumption exist only in households with adult females. Households with increasing numbers of adult males show increases in fresh vegetable expenditures; however, the number of persons under 19 years and above 64 years did not appear to have much influence on expenditures. Race, education and food stamp participation were also found to be insignificant. In contrast, expenditures were significantly positively related to the degree of population density; and with regard to region, households located in the West spend more on fresh vegetables than households residing in the Northcentral and Southern regions, while households located in the Northeast spend more on fresh vegetables than Western households.

Two other studies (Smallwood and Blaylock, 1984; Blaylock and Smallwood, 1986) employed the Tobit model to quantify the impact of economic and demographic variables on household's consumption of vegetables. The 1984 study used the U.S. Department of Agriculture's 1977-1978 Nation-wide Food Consumption Survey, while the other utilized data from the 1980-1981 Continuing Consumer Expenditure Survey. The studies had the following independent variables in common; income, family size, region, race, season, and age composition. Income was a significant determinant of fresh vegetable expenditures. The income elasticity was estimated at 0.15 in the first study and 0.24 in the second. With regard to household size, the studies produced conflicting results--while the 1984 study estimated a significant negative relationship between household size and fresh vegetable expenditures, a significant but positive relationship was established in the 1986 study. Their suggested impact of regional location on expenditures was quite comparable. The first study ranks regions, in order of decreasing tendency to spend on fresh vegetables, as follows: Northeast, West, South and Northcentral. In comparison, the ranking for the second study was in the following order; West, Northeast, South, Northcentral. A result from both studies was that fresh vegetable expenditures are highest in the spring and lowest in the fall, and expenditures in the summer are higher than in the winter. The 1986 study reported that blacks tend to spend less on fresh vegetables than other races as a group. In contrast, a conclusion from the 1984 study was that whites are less likely to spend on fresh vegetables than both blacks and nonwhite/nonblacks.

Most of the above studies reported a positive but inelastic income elasticity for fresh vegetables. Intuitively, however, one would expect fresh-winter vegetables to exhibit greater income elasticities. This intuition is based on the fact that the winter season precludes the production of home grown or commercial vegetables in most states and thus gives rise to higher prices of fresh-winter vegetables. The bunching of winter vegetables and other fresh vegetables into one category of fresh vegetables provides a possible explanation for the inelastic income elasticities reported in other studies.

Among the studies examined in this section, only one (Buse and Salathe) used equivalent scales to incorporate household composition effects in the analysis of vegetable consumption. With regard to unpooled estimation, only Salathe (1979a) employed this method. The remaining studies employed the more convenient approach of simply including these variables on the right hand side of the demand specification.

Censored-Regression Models

All the above studies relating socioeconomic variables to vegetable consumption used cross-sectional data on individual households' or aggregate expenditures on, or quantities consumed of, various vegetable items. With data from individual households, one can expect a significant portion of the observations on the dependent variable to be zero. These observations indicate that some households (in the absence of misreporting) did not purchase the commodity in question during the survey period. Samples for which values of the dependent variable, corresponding to a known set of explanatory variables, can only be observed for a limited range are said to be censored. Models which are based on such samples present special problems of specification and estimation. To illustrate, let y_i^* represent the desired expenditure of the ith house on the commodity in question; then assume that y_{i*} is a linear combination of the explanatory variables and an error term ($y^* = x_i\beta + e_i$). The regression function based on the positive observations of the dependent variable can then be written as

$$E(y_i \mid x_i, y_i > 0) = x_i\beta + E(e_i \mid e_i > - x_i\beta) \qquad (1)$$

The conditional expectation of the error term is generally non-zero, therefore an ordinary least squares regression on the positive observations will provide biased estimates of β (Maddala 1983, pg. 2).

Furthermore, Greene (1981) has shown that the ordinary least squares estimator of β when all the observations on y_i are used is biased and inconsistent. To obtain consistent estimates of the censored regression model, a different method of estimation is required.

Maddala (1983) and Amemiya (1984) have provided excellent reviews of the literature pertaining to the specification and estimation of models that fall within the limited dependent variable framework, so a full review of the literature is not needed here. However, some of the major developments of relevance to the present study are highlighted.

Tobin (1958) analyzed households' expenditures on durable goods and provided the first application of regression methods to censored data. His model, commonly called the Tobit model, is specified as follows:

$$\begin{aligned} y_i^* &= x_i\beta + e_i \qquad && e_i \sim IN(0, \sigma^2) \\ y_i &= yi^* && \text{if } y_i^* > y_0 \\ &= 0 && \text{otherwise} \end{aligned} \qquad (2)$$

where y_i is the ith individual household's expenditure, y_i^* is the desired but unobserved consumption level of that household and x_i represents a vector of explanatory variables that characterizes the households desire to consume the good. In this model $y_i = 0$ because values of y_i^* less than zero are not observed. Thus, as Maddala has pointed out, Tobin use of the model to analyze expenditures on automobiles was inappropriate, because zero observations was an outcome of consumers' choice rather than unobservability. To obtain consistent estimates of β and σ^2, Tobin used the maximum likelihood (ML) estimator. The ML estimator applied to the Tobit model has been shown (Amemiya, 1973) to be consistent and asymptotically normal.

The Tobit model has been widely applied to censored data, however, as was first pointed out by Cragg (1971) the model in certain cases may be an invalid representation of the censoring process. According to the model, the decision on whether to purchase, $Pr(y_i = 0)$, and on how much to purchase, $Pr(y_i > 0)\phi(y_i \mid y_i > 0)$, are based on the same stochastic process (the same variables and parameters). Consequently zero observations on expenditures always imply that the desired or optimal level of consumption, determined via the utility maximization process, is non-positive. Several studies (Cragg(1971), Deaton and Irish(1984), Blundell and Meghir(1987) among others), however, have recognize other possible explanations for zero observations on the dependent variable.

Specifically, the literature has noted two other explanations for the existence of zero observations. The first situation which was initially modeled by Cragg (1971), is one in which the consumer desires a positive amount of the good in question, but purchasing the item depends not only on the intensity of that desire but also on such factors as the availability of the good, amount of search, and the information and transaction cost involved in acquiring the good. Therefore, according to Cragg, for a purchase to occur two hurdles have to be overcome. First, the consumer has to decide whether to purchase and second decide on the amount to purchase. The first decision is closely linked to the desire for the good, while the second to the impediments to purchasing. It is possible, therefore, for the consumer to desire a positive amount, but because the barriers to purchasing are so great no acquisition occurs.

Misreporting by either the respondent or the enumerator and infrequent purchases provide other explanations for zero expenditures. In the case of infrequent purchases, zero expenditures may have been recorded because the consumer did the purchasing before or after the survey period, thus the occurrence of zero expenditures do not necessarily imply that the consumer does not purchase the item nor that the consumer did not consume the item during the survey period.

In summary, a zero observation on the dependent variable could occur because (1) the good is not consumed, (2) a positive amount is desired but certain impediments prohibit purchases, and (3) of misreporting and/or the good was purchase outside of the survey period. As indicated above, the Tobit model captures only the first among these three censoring rules, therefore, if the other censoring rules are present, the Tobit model represents a misspecification.

Although Cragg's Double-Hurdle model and variations of it have been in use since its inception in 1971, models that account for misreporting and/or infrequent purchases are of more recent vintage. Deaton and Irish (1984) Kay, Keen and Morris (1984), Keen (1986), and Blundell and Meghir (1987), are among the pioneers of misreporting/purchasing infrequency models.

Along with the development of alternative specifications to the Tobit model, statistical tests have been constructed specifically to test the Tobit model against these alternatives. Lin and Schmidt (1983) adopted the Langrange Multiplier (LM) test to derive a test for the Tobit model against Cragg's Double-Hurdle model. The test is attractive because only the Tobit model need be estimated. Haines et al. (1988) have since applied the test to adult women's consumption of ten food groups. The hypothesis that the Tobit model was correctly specified against the alternative--Cragg's model--was rejected for nine of those food groups.

Based on a generalized model that nest both the Tobit and Cragg's model, Lee and Maddala (1985) developed LM tests to select the most appropriate specification. They suggest that their LM test statistic of the Tobit model against Cragg's alternative is asymptotically equivalent to Nelson's Hausman test statistic, hence their test statistic can also be used as a general misspecification test.

Problems of heteroscedasticity and non-normality are two other specification considerations of importance to the Tobit model. Unlike the standard regression model, ether heteroscedasticity or non-normality can render maximum likelihood parameter estimates inconsistent (Hurd 1979, Goldberger 1983). Nelson (1981) developed a Hausman test that can be used to test the Tobit model against general misspecifications, including heteroscedasticity and non-normality. For certain population moments, he suggested the maximum likelihood estimator for comparison with the method of moments estimator. The maximum likelihood estimator is both consistent and efficient in the absence of misspecification but inconsistent otherwise, while the method of moments estimator is considered to be consistent both in the presence and absence of misspecifications. To illustrate the test, Nelson chose the likelihood equation associated with the population moment, X'Y. The corresponding test statistic distributed asymptotically as a chi-square with k degrees of freedom was specified as

$$m = N(N^{-1}X'Y - \hat{E}_{xy})'(\hat{V}_1 - \hat{V}_0)^{-1}(N^{-1}X'Y - \hat{E}_{xy}) \tag{3}$$

where $N^{-1}X'Y$ is the method of moment estimator for $E(N^{-1}X'Y)$, E_{xy} the corresponding maximum likelihood estimator, V_1 the variance of $N^{-1}X'Y$ and V_0 is the variance of E_{xy}. This test is computationally burdensome; in addition to obtaining maximum likelihood estimates of β and σ, V_1 involves the first and second moments around y_i and V_0 involves, among other terms, the information matrix $(I(\beta,\sigma)^{-1})$.

Another Hausman test for heteroscedasticity and non-normality in the Tobit model was derived by Newey (1987). He based his test on the difference between Powell's (1986) symmetrically censored least squares (SCLS) estimator and the Tobit maximum likelihood estimator. The SCLS estimator is based on the restriction that the conditional distribution of the regression disturbance term is symmetric around zero. The estimator is thought to be robust to a wide range of non-normal or heteroscedastic disturbance distributions. Newey's specification of the Hausman test statistic is given below.

$$h = n(\hat{\delta}_s - \hat{\delta})'[V(\hat{\delta}_s - \hat{\delta})]^{-1}(\hat{\delta}_s - \hat{\delta}) \tag{4}$$

Where δ_s and δ are, respectively, the SCLS and the Tobit maximum likelihood estimates of β and σ, and $V(\delta_s - \delta)$ is a consistent estimator of the asymptotic covariance matrix of sqrt(n)(δ_s - δ). Like the previous test, this test is difficult to implement - two different set of parameter estimates and covariance matrices are needed in order construct the test. Furthermore, the use of $V_s(\delta)$ - $V(\delta)$ as an estimate of $V(\delta_s - \delta)$ may not always be feasible because the possibility of a negative value for h is not ruled out; thus, if that is the case, an alternative estimator would be required.

To test for heteroscedasticity in the Tobit model, Lee and Maddala (1985) suggested specifying the variance of the heteroscedastic disturbance term as a function of a constant term and a vector of exogenous variables without a constant term. Testing for heteroscedasticity is then reduced to testing whether the coefficient associated with the exogenous variables in the variance term is significantly different from zero. Along those lines, they constructed a LM test which they argue is invariant to the functional form adopted for the heteroscedastic variance structure.

White (1982) has developed an information matrix misspecification test applicable to a wide variety of models, including limited dependent variable models. The test is based on the information matrix equivalent theorem which says that when the model is correctly specified the information matrix can be expressed either as the negative of the Hessian or the outer product of the first derivatives. Thus if the model is misspecified the sum of the two terms is different from zero. Chesher (1984) has derived White's information matrix test as a result of constructing a test for parameter heterogeneity . His findings suggest that the information matrix test is a valuable diagnostic tool for analysts using cross-sectional data to estimate models of individual behavior. In addition Chesher has shown that the variance of the sum of the Hessian and the product of the first derivatives can be obtained without third derivatives of the log likelihood function. In fact the test can be constructed from an artificial regression and requires only the first and second derivatives of the log-likelihood function.

The general specification test by White and Chesher has several attractive features. The test is effective against both parameter inconsistences and distributional assumptions. Also, unlike the Lee and Maddala (1985) LM test for Heteroskedasticity, the information matrix test does not assume knowledge of the disturbance structure. Furthermore, the test is based on just the maximum likelihood estimator and thus does not require a second estimator as in the case of the Hausman tests developed by Nelson(1981) and Newey (1987).

Although a number of tests for heteroskedasticity and non-normality have been developed for the Tobit model, few corrective measures have been suggested. In the case of heteroskedasticity, Maddala (1983) has indicated that it is more practical to make some reasonable assumption about the nature of heteroskedasticity and estimate the model than to ignore the problem altogether. Fishe et al. (1979) and later Bomberger and Denslow (1980) estimated the Tobit model with the variance of the error term specified as a function of a constant term and a subset of the independent variables.

According to Maddala (1983) there are two ways of treating non-normal errors: (1) devise methods of estimation for non-normal distributions or (2) use transformations to normality. Amemiya and Boskin (1974) have considered the estimation of a censored regression model with a log-normal distribution. Maddala (1983) has suggested the exponential or gamma distribution as alternative error distributions in the context of censored or truncated regression models. A disadvantage of this first method of dealing with non-normality is that it assumes a priori knowledge of the form of the non-normal distribution.

The Box-Cox transformation is commonly used as a transformation to normality. However, Maddala (1983) has pointed out that because the transformation imposes restrictions on the range of the transformed error terms the assumption of normality is not valid. Rather, the residuals should be considered truncated normally distributed. Poirier (1978) has considered estimation in the case where the error terms are assumed to have a truncated normal distribution.

The inverse hyperbolic sine (IHS) transformation apparently has not been applied to limited dependent variable models or used in demand analysis. Burbidge et al. (1988) were one of the few users of this transformation. However, this transformation holds much promise as a transformation to normality. The transformation is continuously defined over positive, zero, and negative values and thus is more likely to produce normally distributed residuals than the Box-Cox transformation.

CHAPTER 3
MODEL SPECIFICATION

This chapter discusses the specification and estimation of alternative censored regression models. Misspecification testing and modifications to account for heteroscedasticity and non-normality are also discussed.

The study is based on cross-sectional data on individual household's expenditures on fresh-winter vegetables. With data from individual households one can expect a number of the observations on the dependent (fresh vegetable expenditures) variable to be zero. As indicated in chapter two, this phenomenon renders ordinary least squares an inappropriate estimator. Three main reasons have been given for the existence of zero expenditures:

1. The good is not desired and hence is not consumed;
2. Impediments such as transaction and information cost, availability of the good, and the amount of search involved in purchasing the good, prohibit purchases;
3. Expenditures were misreported, or because the good is purchased infrequently, a discrepancy exists between observed purchases and unobserved consumption.

Each of those reasons for the occurrence of zero expenditures are associated with a different censored-regression model or model specification.

The Tobit Model

The Tobit model as developed by Tobin (1958) embodies the first of these censoring rules (explanations for the occurrence of zero expenditures) and is specified as follows:

$$\begin{aligned} y_i^* &= x_i\beta + e_i && e_i \sim IN(0,\ \sigma^2) \\ y_i &= y_i^* && \text{if } y_i^* > 0 \\ &= 0 && \text{otherwise} \end{aligned} \tag{5}$$

where y_i is the ith individual household's observed expenditure on fresh-winter vegetables, y_i^* is the desired or optimal expenditure level of that household and can be construed as the solution to a utility maximization problem, and x_i represents a vector of explanatory variables (namely socio-economic and demographic variables) that characterizes the household's preferences. According to this specification, observed expenditures is equal to the desired expenditure level if desired expenditures is greater than zero; otherwise zero expenditures are observed. Desired expenditure, y_i^*, can take on negative values. However, values of y^* less than zero are unobserved, hence, y_i is censored at zero.

Equation 5 implies that the probability of zero observations ($y_i=0$) is

$$\begin{aligned} Pr(y_i = 0) = Pr(y_i^* < 0) &= 1 - Pr(y_i^* > 0) \\ &= 1 - \Phi_i(x_i\beta/\sigma) = 1 - \Phi_i(e) \end{aligned} \tag{6}$$

where $\Phi_i(e)$ is the standard normal distribution function evaluated at $x_i\beta/\sigma$. With regard to the positive observations ($y_i > 0$), we have

$$= \Pr(y_i^* > 0)f(y_i \mid y_i^* > 0)$$

$$= \Pr(y_i^* > 0)f(y_i - x_i\beta, \sigma^2) / \Pr(y_i^* > 0)$$

$$= f(y_i - x_i\beta, \sigma^2) = 1/(2\pi\sigma^2)^{½} \, \mathrm{EXP}(-(1/2\sigma^2)(y_i - x_i\beta)^{-}) \qquad (7)$$

The log likelihood for the Tobit model is thus

$$\text{Log } L = \sum_0 \log(1 - \Phi_i(e)) - (N_1/2)\log 2\pi - (N_1/2)\log \sigma^2$$

$$- \sum_1 (1/2\sigma^2)(y_i - x_i\beta)^2 \qquad (8)$$

where $\sum_0$ and $\sum_1$ refer to summation over zero and positive observations, respectively, and N_1 indexes the observations associated with the positive values of y. The first derivatives of the log likelihood function follows as

$$\frac{\partial \log L}{\partial \log \beta} = - \frac{1}{\sigma} \sum_0 \frac{x_i\phi_i}{1 - \Phi i} + \frac{1}{\sigma^2} \sum_1 (y_i - x_i\beta)x_i \qquad (9)$$

$$\frac{\partial \log L}{\partial \log \sigma^2} = \frac{1}{2\sigma^3} \sum_0 \frac{x_i\beta_i\phi_i}{1 - \Phi i} - \frac{N_1}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_1 (y_i - x_i\beta)^2 \qquad (10)$$

With these first derivatives maximum likelihood estimates of σ^2 and β can be obtained via the method of Berndt, Hall, Hall, and Hausman. Alternatively, the method of Newton which uses the first and second derivatives can be utilized. The maximum likelihood method applied to the Tobit model, under the distributional assumptions of homoscedasticity and normality, has been shown (Amemiya, 1973) to be consistent and asymptotically normal.

The Double-Hurdle Model

Cragg's Double-Hurdle model generalizes the Tobit model in that it recognizes that, although the household may desire a positive amount of the good, impediments to acquisition may prohibit purchases. This recognition led to the modelling of consumption behavior in two stages: first, based on the impediments to acquisition the household decides whether or not to purchase the good, and second, according to the intensity of the desire for the good the household decides on how much to purchase. The Double Hurdle model is represented as

$$\begin{aligned} y_i = y_i^* \quad & D_i > 0 \\ 0 \quad & \text{otherwise} \end{aligned} \tag{11}$$

$$\begin{aligned} D_i &= z_i\theta + v_i \\ y_{i*} &= x_i\beta + e_i \end{aligned} \tag{12}$$

where y_i and y_i^* are previously defined, and D_i characterizes the decision of whether to purchase. It is assumed that only the sign of D_i is observed and that y_i^* is observed only when D_i is positive. The vectors of independent variables (x_i, z_i) need not be different, and the error terms (e_i, v_i) are assume to be independently normally distributed with zero means and constant variances (σ^2, 1). According to the above specification, before purchases are realized the household must surpass the first hurdle of deciding whether or not to purchase and the second which involves the decision of how much to purchase-- hence the term double hurdle. This specification pinpoints the essential difference between the Tobit and Double-Hurdle model. In the Tobit model the same variables (x_i) and parameters (β_i) explain the decision on whether to purchase and on how much to purchase, in contrast, the Double-Hurdle model allows different sets of variables and parameters (x_i, z_i; β_i, θ) to characterize the two decisions. However, the double hurdle model does not preclude the possibility that the two sets of variables and parameters are identical.

According to the Double-Hurdle model the probability of zero observations ($y_i = 0$) is

$$\begin{aligned} Pr(y_i = 0) &= Pr(D_i < 0) \\ &= 1 - Pr(D_i > 0) \\ &= \underline{1 - \Phi_i(z_i\theta) = 1 - \Phi_i(v), \quad \text{where } v = z_i\theta} \end{aligned} \tag{13}$$

With regard to the positive observations ($y_i > 0$) we have

$$\begin{aligned} &Pr(D_i > 0)f(y_i \mid y_i^* > 0) \\ &= Pr(D_i > 0)f(y_i) \,/\, Pr(y_i^* > 0) \quad \text{conditional} \\ &= \Phi_i(v)/\Phi_i(e)f((y_i - x_i\beta, \sigma^2). \end{aligned} \tag{14}$$

The log-likelihood function for the Double-Hurdle model is thus

$$\log L = \sum_0 \log(1 - \Phi(v)) + \sum_1 (\log \Phi(v) - \log \Phi(e)) - (N_1/2)\log 2\pi$$

$$- (N_1/2)\log \sigma^2 - \sum_1 (1/2\sigma^2)(y_i - x_i\beta)^2 \tag{15}$$

The corresponding first derivatives are

$$\frac{\partial \log L}{\partial \log \beta} = - \frac{1}{\sigma} \sum_1 \frac{x_i\phi_i}{\Phi i} + \frac{1}{\sigma^2} \sum_1 (y_i - x_i\beta)x_i \tag{16}$$

$$\frac{\partial \log L}{\partial \log \theta} = - \sum_0 \frac{z_i\phi_i(v)}{1 - \Phi i(v)} + \sum_1 \frac{z_i\phi_i(v)}{\Phi i(v)} \tag{17}$$

$$\frac{\partial \log L}{\partial \log \sigma^2} = \frac{1}{2\sigma^3} \sum_0 \frac{x_i\beta\phi_i}{\Phi i} - \frac{N_1}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_1 (y_i - x_i\beta)^2 \tag{18}$$

Given the log-likelihood function (log L) and its associated derivatives, maximum likelihood estimates of the parameters, (β, θ, σ), can be obtained in a similar fashion as in the Tobit model. Ordinary least squares estimates can be used as starting values for β and σ, while starting values for θ can be obtained from estimates of the Probit model.

Recognizing that when $\Phi(v) = \Phi(e)$ the Double-Hurdle model is reduced to the Tobit model (thus the Tobit model is nested in the Double Hurdle model), the Likelihood Ratio test statistic or some form of a score statistic can be used to test the Double-Hurdle against the Tobit specification.

Purchase Infrequency Model

In analyzing consumer behavior, the variable of interest is usually consumption levels and not expenditures per se. However, the data at hand contains expenditures on fresh-winter vegetables, rather than amounts actual consumed during the survey period. To the extent that observed expenditures are identical to consumption levels, no inconsistencies exist with the use of expenditure data. As alluded to above, discrepancies between observed expenditures and unobserved consumption are likely to exist if the good is purchased infrequently. In fact, the occurrence of zero expenditures may result from infrequent purchases. Both the Tobit and Double-Hurdle model assume that observed positive expenditures are identical to the unobserved consumption level. Thus, if discrepancies exist between expenditures and consumption levels, the Tobit and Double-Hurdle model will be inconsistent with the data generating process or the underlying consumption behavior. Given that the data for the study are comprised of expenditures over a two week period and fresh vegetables are not likely to be stored for over two weeks, such discrepancies are not expected to be a serious factor. However, it may be useful to compare the results of the Purchase Infrequency model with that of the Tobit and Double-Hurdle model, as a casual test of the hypothesis that discrepancies do not exist between observed expenditures on, and the consumption of, fresh-winter vegetables.

The Purchase Infrequency model adopted from Blundell and Meghir (1987) assumes that positive amounts of the good are always consumed but because the good is purchased infrequently, expenditures may not always correspond with consumption, hence the realization of zero expenditures during the survey period. This discrepancy that may exist between observed expenditures and unobserved consumption was used to motivate the following Purchase Infrequency model specification.

As before, let y_i be observed expenditures, and y_i^* ($=x_i\beta+e_i$) is the unobserved consumption level. Also, let D_i ($= z_i\theta + v_i$) be an unobserved variable characterizing the purchase infrequency phenomenon. The error term, v_i, is normally distributed with zero means and constant variance equal to one. It is assumed that $D_i > 0$, if and only if $y_i > 0$. Now assuming that the expected value of expenditures, $E(y_i)$, is equal to the expected value of consumption, $E(y_i^*)$, we have

$$E(y_i) = Pr(y_i > 0)E(y_i \mid y_i > 0) + Pr(y_i = 0)E(y_i \mid y_i = 0)$$

Since $D_i > 0$ if and only if $y_i > 0$, the above expression can be written as

$$\begin{aligned} E(y_i) &= Pr(D_i > 0)E(y_i \mid D_i > 0) + Pr(y_i = 0)E(y_i \mid D_i \le 0) \\ &= Pr(D_i > 0)E(y_i \mid D_i > 0) \end{aligned}$$

$$\text{Thus} \quad \Phi(v)y_i = E(y_{i*}) \tag{19}$$

Letting an error term w_i represent discrepancies due to infrequent purchases and/or misreporting, between observe expenditures and actual consumption, (19) can be written as

$$\Phi(v)y_i = y_i^* + w_i = (x_i\beta + e_i) + w_i \tag{20}$$

were both e_i and w_i are assumed to have zero means and constant variances. The infrequency purchase model can thus be specified as

$$y_i = (y_i^* + w_i) \,/\, \Phi(v) \qquad D_i > 0$$
$$= 0 \qquad \text{otherwise} \tag{21}$$

Allowing $u_i = e_i + w_i$, and assuming that u_i is independent of v_i, the model can also be specified as

$$\Phi(v)y_i = x_i\beta + u_i \qquad y_i > 0$$
$$= 0 \qquad \text{otherwise} \tag{22}$$

From this specification, the contribution of the zero observations to the likelihood function is identical to that of the Double-Hurdle model

$$\Pr(y_i = 0) = 1 - \Phi(v) \tag{23}$$

and for the positive observations, we have

$$\Pr(y_i > 0)f(y_i \mid y_i > 0)$$
$$= \Pr(D_i > 0)f(y_i \mid y_i^* > 0)$$
$$= \Pr(D_i > 0)f(y_i) \,/\, \Pr(y_i^* > 0)$$
$$= \Pr(D_i > 0)f(y_i) = \Phi(v)f(y_i - x_i\beta,\ \sigma^2)$$
$$= \Phi(v)|J|(1/\sigma)\phi(\ (\Phi(v)y_i - x_i\beta)/\sigma) \tag{24}$$

where $J = \Phi(v)$ is the Jacobian term in (22).

The log-likelihood function for the Purchase Infrequency model follows as

$$\log L = \sum_0 \log(1 - \Phi(v)) + 2\sum_1 \log \Phi(v) - (N_1/2)\log \sigma^2$$

$$- (N1/2)\log 2\pi - \sum_1 1/(2\sigma^2)(\Phi(v)y_i - x_i\beta)^2 \qquad (25)$$

and the first derivatives are as follows

$$\frac{\partial \log L}{\partial \log \beta} = \frac{1}{\sigma^2}\sum_1 (\Phi_i(v)y_i - x_i\beta)x_i \qquad (26)$$

$$\frac{\partial \log L}{\partial \log \theta} = -\sum_0 \frac{\phi_i(v)z_i}{1 - \Phi_i(v)} - \frac{1}{\sigma 2}\sum_1 (\Phi_i(v)y_i - x_i\beta)\phi_i(v)z_iy_i$$

$$\sum_1 \frac{2\phi_i(v)z_i}{1 - \Phi_i(v)} \qquad (27)$$

$$\frac{\partial \log L}{\partial \log \sigma^2} = -\frac{N1}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_1 (\Phi_i(v)y_i - x_i\beta)^2 \qquad (28)$$

As in the case of the Double-Hurdle model, the first and second derivatives of the log-likelihood function with respect to $\delta=(\beta,\theta,\sigma)$ can be used to obtain maximum likelihood estimates of δ.

Misspecification and Transformation of Censored Regression Models

Unlike standard regression models, if the data exibits non-normality, maximum likelihood estimates of the Tobit and Double-Hurdle model would be inconsistent. Similarly, the presence of heteroscedasticity would render maximum likelihood estimates of all three of the models presented above inconsistent. Thus such misspecification is of particular importance to censored regression models. This section presents a systematic approach of testing and respecification to account for heteroscedasticity and non-normality in censored regression models. Rather than repeating the suggested procedure for each model, the Tobit model will be used for illustrative purposes. White's information matrix test will be employed to detect misspecifications. Maddala's suggested treatment will be used to accommodate heteroscedasticity, while the inverse hyperbolic sine transformation (IHS) will be employed as a transformation to normality.

Each of the models specified above represents a different way of mapping unobservable consumption levels y_i^* ($= x_i\beta + e_i$) to the observable counterpart y_i, depending on their conceptualization of the occurrence of zero expenditures. The unobservability of y_i^*, however, precludes the estimation of the residuals (e_i). Consequently, familiar residual based tests useful for inferring serial correlation, heteroscedasticity and non-normality in standard regression models are not directly appropriate (see Grourieroux et al., 1987). White's information matrix test which is based on maximum likelihood estimates provides an operational alternative.

The test is based on the information matrix equivalence theorem which implies that when the model is correctly specified, the information matrix can be expressed either as the Hessian of the log likelihood function or the outer product of its first derivatives. Accordingly, the following equality

$$A(\delta) = -\Sigma \frac{\partial^2 \log L}{\partial\delta_i \partial\delta_j} = \Sigma \frac{\partial \log L}{\partial\delta_i} \quad \Sigma \frac{\partial \log L}{\partial\delta_j} = B(\delta) \qquad (29)$$

where for the Tobit model $\delta=(\beta,\sigma^2)$, should hold. Equivalently, equation 29 can be expressed as

$$D(\delta) = \Sigma \frac{\partial^2 \log L}{\partial\delta_i \partial\delta_j} + \Sigma \frac{\partial \log L}{\partial\delta_i} \quad \Sigma \frac{\partial \log L}{\partial\delta_j} = 0 \qquad (30)$$

Interpreting large deviations of $D(\delta)$ from zero as evidence of misspecification, White's information matrix test statistic, distributed as a chi-square, is constructed as

$$In = D(\hat{\delta}) \; V(\hat{\delta})^{-1} \; D(\hat{\delta})' \qquad (31)$$

where $\hat{\delta}$ is the maximum likelihood estimate of $\hat{\delta}$, and $V(\delta)^{-1}$ is the covariance matrix of $D(\hat{\delta})$. White's formulation of $V(\hat{\delta})^{-1}$ may pose computational difficulties because it involves third derivatives of the log likelihood function. Fortunately, however, Chesher has developed a construction of the Information matrix test that requires only the first and second derivatives. The test statistic was shown to be n times the R^2 from the least squares estimation in which a column vector of ones is regressed on a matrix with elements

$$\frac{\partial \log L}{\partial\delta_j} \quad \text{and} \quad \frac{\partial^2 \log L}{\partial\delta_i \partial\delta_j} + \frac{\partial \log L}{\partial\delta_i} \quad \frac{\partial \log L}{\partial\delta} \qquad (32)$$

The information matrix test for the parameter vector β in the case of the Tobit model is illustrated below

$$A(\beta) = \Sigma \frac{\partial^2 \log L_i}{\partial\beta^2} = -\sum_0 \frac{f_i{}^2 x_i{}^2}{(1-F_i)^2} + \sum_0 \frac{\beta x_i 2 f x_i}{\sigma^2(1-F_i)} - \sum_1 \frac{x_i{}^2}{\sigma^2} \tag{33}$$

$$B(\beta) = \Sigma \frac{\partial^2 \log L_i}{\partial\beta} \frac{\partial \log L_i}{\partial\beta'} = \sum_0 \frac{f_i{}^2 x_i{}^2}{(1-F_i)^2} + \sum_1 \frac{(y - x_i\beta)2x_i{}^2}{\sigma^4} \tag{34}$$

where $f_i = \phi(e)\sigma^{-1}$, and $F_i = \Phi(e)$. Clearly, the first terms on the RHS of $A(\beta)$ and $B(\beta)$ cancel, however, if there is heteroscedasticity the second term on the RHS of $B(\beta)$ will be too large relative to the remainder of $A(\beta)$ because large squared deviations of $y_i - x_i\beta$ will be associated with large x_i's. In a similar manner the relationships between $\partial^2\log L/(\partial\sigma^2)^2$ and $(\partial\log L/\partial\sigma^2)^2$, and between $\partial^2\log L/\partial\sigma^2\partial\beta$ and $\partial\log L/\partial\sigma^2 \cdot \partial\log L/\partial\beta$ can be used to indicate kurtosis and skewness.

If upon application of the information matrix test the null hypothesis of no misspecification is rejected, a likely suspect is non-normality. As mentioned in the literature review, there are two general approaches for dealing with non-normal disturbances--transformation of the data or imposing a different distribution. This latter approach is not particularly attractive because it presumes knowledge of what non-normal distribution is actually generating the errors. Considering transformations to normality, the only technique used to date appears to be the Box-Cox transformation.

Denoting the Box-Cox transformation as T (.), then its use in equation 5 would imply

$$T(y_i^*) = x_i\beta + e_i^* \qquad (35)$$

with e_i^* now normally distributed. Maddala has pointed out, however, that e_i^* cannot be normally distributed since T(.) is not defined for $y_i^* < 0$. Thus a truncated normal distribution must be assumed. As an alternative, the inverse hyperbolic sine transformation, I(.), is continuously defined over positive, zero, and negative values. The transformation yields

$$I(y_i^*) = x_i\beta + e_i^* \quad \text{and}$$

$$I(y_i) = I(y_i^*) \quad y_i^* > 0$$

$$I(y_i) = y_i = 0, \text{ otherwise.} \qquad (36)$$

The specific form of the transformation is

$$I(y_i) = \log(\alpha y_i + (\alpha^2 y_i^2 + 1)^{\frac{1}{2}})/\alpha \qquad (37)$$

where α is a scaler location parameter that can be estimated from the data. As pointed out by Burbidge et al. I(.) is symmetric about zero in

α, the limit of $I(y_i)$ as α goes to zero is y_i and for relatively large values of α the transformation behaves logarithmically. Use of the IHS transformation changes the log-likelihood function of equation 8 in two ways. First $I(y_i)$ replaces y_i and a term accounting for the jacobian of the transformation is included to yield

$$\text{Log } L = \sum_0 \log(1 - \Phi_i(e)) - (N_1/2)\log 2\pi - (N_1/2)\log \sigma^2 + \sum_1 (1/2\sigma^2)(I(y_i) - x_i\beta)^2 - \tfrac{1}{2} \sum_1 \log(1 + \alpha^2 y_i^2) \quad (38)$$

Although the resulting log likelihood of the Tobit model is highly non-linear in α, an estimation strategy which first determines the maximum likelihood estimates of β conditioned on the specification of equation 8 and then uses these as starting values in equation 39 should be successful if the initial value of α is set close to zero. Following the non-normality fix-up, if the null hypothesis is again rejected, the next likely suspect is heteroscedasticity, because cross sectional data are predisposed to exhibiting heteroscedasticity. For example, in the present study, very large households and/or households with unusually high incomes are likely to exhibit considerably more variability in the consumption of fresh-winter vegetables than the average household. To correct for heteroscedasticity, Maddala suggests modelling the variance as a function of a constant and exogenous variables expected to be related to the variance. For the study at hand, household income, household size and composition are likely candidates. Thus, specifying the variance as

$$\sigma_i^2 = f(Z_i, \tau) \tag{39}$$

where Z is a subset of the exogenous variables, and τ is a vector of parameters to be determined, the log-likelihood function for the Tobit model in the presence of the IHS transformation becomes

$$\text{Log } L = \sum_0 \log(1 - \Phi_i(e)) - (N_1/2)\log 2\pi - (N_1/2)\log \sigma_i^2$$

$$+ \sum_1 (1/2\sigma_i^2)(I(y_i) - x_i\beta)^2 - \tfrac{1}{2} \sum_1 \log(1 + \alpha^2 y_i^2) \tag{40}$$

Interpretation and Predictions

The Censored Regression models specified above can be used to obtain predictions of fresh-winter vegetables by deriving various expectation functions. Three different predictions are illustrated below--desired but unobserved expenditures, expenditures conditional on the information that expenditures are greater than zero, and unconditional expenditures. Because of the IHS transformation, these expected values differ from that of the traditional Tobit model. For desired unobserved expenditures we have

$$E[I(y_i^*)] = x_i\beta \tag{41}$$

However, what is sought for is $E(y^*)$. Recognizing that $I(y) = \sinh^{-1}(y)/\alpha$, the result that $\sinh(y) = (e^y - e^{-y})/2$ is used to obtain $\text{Plim}(y_i^*)$ as

$$\text{Plim}(y_i^*) = [\exp(\alpha x_i\beta) - \exp(-\alpha x_i\beta)]/2\alpha \tag{42}$$

Because of the exponential involved in obtaining the hyperbolic sine function, Plim(.) is used instead of E(.). For unconditional expenditures we have

$$\begin{aligned} E[I(y_i)] &= Pr(I(y_i)\ E[I(y_i)\ |\ I(y_i) > 0] \\ &= Pr(I(y_i)E[I(y_i)\ |\ e_i > -x_i\beta] \\ &= \Phi(e)\ (x_i\beta + E[e_i\ |\ e_i > -\ x_i\beta]) \\ &= \Phi(e)\ (x_i\beta + \sigma_i\phi(e)/\Phi(e)) \\ &= \Phi(e)x_i\beta + \sigma_i\phi(e) \end{aligned} \tag{43}$$

Thus, in a similar fashion as above, $Plim(y_i)$ is derived as

$$Plim(y_i) = [\exp(\alpha\Phi(e)x_i\beta + \alpha\sigma_i\phi(e)) - \exp(-\alpha\Phi(e)x_i\beta - \alpha\sigma_i\phi(e))]/2\alpha \tag{44}$$

Finally, for the conditional expenditures, we have

$$\begin{aligned} E(I(y_i)\ |\ I(y_i) > 0) &= E[I(y_i)\ |\ e_i > -\ x_i\beta] \\ &= x_i\beta + E(e_i\ |\ e_i > -\ x_i\beta) \\ &= x_i\beta + \sigma_i\phi_i(e)/\Phi_i(e) \end{aligned} \tag{45}$$

Thus

$$\begin{aligned} Plim(y_i\ |\ y_i > 0) = [&\exp(\alpha x_i\beta + \alpha\sigma_i\phi_i(e)/\Phi_i(e)) \\ &- \exp(-\ \alpha x_i\beta\ -\ \alpha\sigma_i\phi_i(e)/\Phi_i(e))]/2\alpha \end{aligned} \tag{46}$$

Given maximum likelihood estimates of the IHS-heteroskedastic-Tobit model and future values of the explanatory variables, predictions based on equations 42, 44 and 46 are obtained. It is also desirable to predict the impact of individual explanatory variables. These are obtained as follows:

$$\frac{\partial \mathrm{Plim}(y^*)}{\partial x_j} = \frac{\beta_j}{2} [\exp(\alpha x_i\beta) + \exp(-\alpha x_i\beta)] \tag{47}$$

$$\frac{\partial \mathrm{Plim}(y)}{\partial x_j} = \frac{\Phi_i(e)\beta_j}{2} [\exp(\alpha\Phi_i x_i\beta + \alpha\sigma_i\phi_i) + \exp(-\alpha\Phi_i x_i\beta - \alpha\sigma_i\phi_i)] \tag{48}$$

$$\frac{\partial \mathrm{Plim}(y \mid y > 0)}{\partial x_j} = \frac{\beta_j}{2} (1 - x_i\beta/\sigma_i \; \phi_i/\Phi_i - (\phi_i/\Phi_i)^2) * [\exp(\alpha x_i\beta + \alpha\sigma_i\phi_i/\Phi_i) + \exp(-\alpha x_i\beta - \alpha\sigma_i\phi_i/\Phi_i] \tag{49}$$

Although the applications illustrated in this chapter pertain to the Tobit model, the Double-Hurdle or Purchase Infrequency model could have been used.

Heckman's Two-Step Estimator

Several consistent estimators have been proposed as approximations to the maximum likelihood estimator associated with censored regression models (the Tobit model in particular). As approximations these estimators, which include Heckman's two-step estimator (Heckman 1976); the Method of Moments estimator (Nelson 1981), the Least absolute Deviations estimator (Powell 1984), and the Symmetrically-Trimmed-Least Squares estimator (Powell 1986), are in general not as efficient as the maximum likelihood estimator provided that the distributional assumptions of the Tobit model holds. Among approximations to the maximum likelihood estimator, Heckman's two-step estimator has probably been used most; thus a comparison of the maximum likelihood estimates with that of Heckman's two-step estimator may be instructional.

The two-step estimator developed by Heckman (1976), followed a suggestion by Gronau (1974). The estimator applied to the Tobit model is illustrated below. Using only the positive observations on y_i

(expenditures on fresh-winter vegetable expenditures), the expected value of y_i can be expressed as

$$E(y_i \mid y_i > 0) = x_i\beta + E(e_i \mid e_i > - x_i\beta), \tag{50}$$

and assuming that the disturbance term, e_i, is normally distributed the above expression can be shown to be

$$E(y_i \mid y_i > 0) = x_i\beta + \sigma\lambda(x_i\beta/\sigma), \tag{51}$$

where $\lambda(.) = \phi(.)/\Phi(.)$. Equation 51 can be rewritten as

$$y_i = x_i\beta + \sigma\lambda(x_i\omega) + \mu_i, \quad \text{for i such that } y_i > 0, \tag{52}$$

where $\omega = \beta/\sigma$, and $\mu_i = y_i - E(y_i \mid y_i > 0)$ such that $E\mu_i = 0$. The variance of μ_i is given as

$$Var(\mu_i) = \sigma^2 - \sigma^2 x_i\omega\lambda(x_i\omega) - \sigma^2\lambda(x_i\omega)^2. \tag{53}$$

Since $Var(\mu_i)$ is a function of the explanatory variables equation 52 is a heteroscedastic regression model. To obtain estimates of β and σ Heckman proposed first estimating ω by the probit maximum likelihood estimator using all the observations on y_i and second regress y_i on x_i and $\lambda(x_i\omega)$ by least squares using only the positive observations on y_i. In vector notation, the second stage of the procedure can be expressed as

$$\hat{\Gamma} = (Z'Z)^{-1}Z'y \tag{54}$$

where $Z = (X, \lambda)$ and $\Gamma = (\hat{\beta},\hat{\sigma})'$. Following Amemiya (1984) and White (1980) consistent estimates of the variance-covariance matrix of Γ can be obtained by

$$(Z'Z)^{-1}Z'\hat{\Omega}Z(Z'Z)^{-1} \tag{55}$$

where $\hat{\Omega}$ is the diagonal matrix whose ith diagonal element is

$[y_i - x_i\hat{\beta} - \hat{\sigma}\lambda(x_i\hat{\omega})]^2$.

Since $\mathrm{Var}(\mu_i)$ is a function of the explanatory variables equation 55 is a heteroscedastic regression model, therefore, more efficient estimates can be obtained by using Generalized Least Squares.

CHAPTER 4
RESULTS AND DISCUSSION

This chapter involves the presentation and analysis of the estimates of the models. In attempting to select the model specification that is most consistent with the data generating process, several statistical tests were employed. The model specification that seems to best fit the data was used to interpret the impact of individual explanatory variables on fresh-winter vegetable consumption, and to obtain corresponding elasticities. In addition the results were used to generate long term forecast of U.S. consumption of fresh-winter vegetables.

The analysis is concerned with fresh vegetables (excluding potatoes) consumed during the months of March, April, May, June, November, and December. The commodity--fresh vegetables--and the specific six months chosen for the analysis represent part of an ongoing research project to study the demand for Florida-fresh-winter vegetables during its major production months.

Although the diary survey which generated the data spans a two week period, 22 percent of all households (a minority) participated during only one week. This discrepancy could be dealt with by deleting those households which participated for only one week. Another approach would be to use weekly expenditures on fresh vegetables, as oppose to biweekly expenditures, as the dependent variable. However, with this second approach, some households would account for two observations while others would account for only one observation. The approach taken here, is to accommodate those households which participated for only one week, by averaging the expenditures of the other households over the two-week period (dividing by two). The resulting sample consisted of 3368 households (observations). Of these, 1088 reported no fresh-winter vegetable expenditures. This significant portion of observations on fresh-winter vegetable expenditures (the dependent variable) taking a zero value provides justification for considering censored-regression models as an appropriate framework for conducting the present investigation.

Other than household income, traditional economic theory generally does not give specific indications of the variables (variables that comprise the vector x_i) to include in the specification of an Engel curve. Consequently, logic, results of past studies, and to a limited extent economic theory, are used to guide the selection of explanatory variables. To begin with, household production theory would suggest that variables characterizing labor market participation (hours of work for example) should influence fresh vegetable consumption. This is expected because labor market participation, in part, reduces the amount of time available to the household for the transformation of fresh vegetables to meal items, thus ultimately constraining the household production function and hence its fresh vegetable expenditures. Household size is another variable that can be expected to influence consumption: apart from the fact that larger households will generally need more food than smaller households, household size introduces economies of scale into consumption. The family life cycle hypothesis provides justification for including household age composition. According to the life cycle concept, biological and psychological changes associated with aging give rise to changing nutritional needs. Thus we can expect the age of household members to influence food consumption patterns. For similar reasons the sex of household members can also be expected to affect food intake. The educational level of the household head can also be expected to influence consumption provided that the level of education affects the dietary choice of the meal planner. Due to differences in tradition, environment, and opportunities (availability of certain goods) associated with location (rural or urban, regions: Northeast, Midwest, South, West), the location of the household is likely to have an impact on its consumption pattern. Varying traditions and consumption habits among races can also influence current and future consumption patterns. The results of past studies (Chapter 2), suggest that most of these variables do impact fresh vegetable consumption.

In chapter 2, ways to incorporate household composition effects into demand equations were discussed. Adult equivalent scales and commodity specific scales are sometimes used to account for differences in consumption arising from such differences as household size, age, and sex. However, equivalent scales introduce additional complexities since their incorporation usually involve the use of specialized functions. The method followed in this study is to simply include the variables on the right side of the Engel curve specification. This approach is ad hoc. However, it avoids the difficulty of using equivalent scales, but, at the same time, allows for differences in fresh-winter vegetable consumption arising from socio-demographic factors.

Table 3 provides a description of the variables included in the analysis. The variables described by if statements were one-zero variables. Averaged weekly household fresh-winter vegetable (exclude potatoes) expenditures was used as the dependent variable. The independent variables include total household food expenditures, household size and household size squared, the age, sex, race, education and marital status of the household head, the age distribution of the household, the region in which the household is located and the months during which the household was surveyed. Obtaining reliable income data on individual households can be quite illusive; for example, some households in the sample did not provide complete information on their income. To circumvent this problem total food expenditure was used in lieu of household income.

Apart from the included explanatory variables, variables such as the number of earners in the household and hours per week the household head worked, designed to characterize the household's labor force participation, were entertained but found to be insignificant. In addition, low order polynomials involving food expenditures, family size and age were considered, but the insignificant coefficients associated with these variables implied that the interactive effect among these variables were minimal.

Model Selection

The results of the Tobit, Double-Hurdle and Purchase Infrequency model are presented in Table 4. Gauss (Edlefsen and Jones), a micro-

Table 3. Variable Definitions

Variable	Mean	Definition
Dependent variable	1.5132	Weekly fresh winter vegetable (excluding potatoes) expenditures (in dollars)
$(\text{Food Expenditure})^{1/2}$	6.2798	Sqrt. of total food at home expenditure (in dollars)
Household Size	2.6113	Number of household occupants
$(\text{Household Size})^2$	6.8189	Household size squared
Age	46.6093	Age of reference person
Sex	0.6698	= 1 if reference person is male
Race		
White	0.8548	Omitted base group
Black	0.1146	= 1 if reference person is black
Nonwhite/nonblack	0.0306	= 1 if reference person is nonwhite/nonblack
Education	0.7289	= 1 if reference person completed H.S.
Marital Status	0.5751	= 1 if reference person is married
Urban	0.8925	= 1 if household resides in urban area
Region		
Northeast	0.3124	Omitted base region
Midwest	0.2360	= 1 if household resides in the MW
South	0.2369	= 1 if household resides in the South
West	0.2147	= 1 if household resides in the West
Season	0.4486	= 1 if household was surveyed during the winter months of November and December
Household Composition		
Children < 5	0.0221	Proportion of household 0-2 yrs old
Children 5 to 13	0.0835	Proportion of household 5-13 yrs old
Persons 14 to 24	0.1866	Proportion of household 14-24 yrs old
Persons 25 to 44	0.3021	Omitted base group
Persons 45 to 64	0.2346	Proportion of household 45-64 yrs old
Persons > 65	0.1712	Proportion of household over 65 yrs old

Table 4. Censored Regression Models of Fresh Winter Vegetable Expenditures.

Variable	Tobit Model	Double-Hurdle Model		Purchase Infreq. Model	
	β_i	β_i	θ_i	β_i	θi
Constant	-3.4623 (0.2761)	-12.5874 (1.0694)	-1.4435 (0.2026)	-1.9648 (0.2835)	-0.9516 (0.1411)
(Food Exp.)$^{1/2}$	0.7401 (0.0192)	1.4037 (0.0529)	0.3874 (0.0136)	0.6006 (0.0150)	0.3488 (0.0113)
Household Size	-0.2676 (0.1204)	-0.5657 (0.3602)	0.0227 (0.0935)	-0.4058 (0.1116)	0.0734 (0.0687)
(Household Size)^2	0.0238 (0.0128)	0.0466 (0.0342)	-0.0106 (0.0114)	0.0435 (0.0104)	-0.0131 (0.0083)
Age	0.0089 (0.0057)	0.0405 (0.0179)	-0.0020 (0.0041)	0.0076 (0.0055)	-0.0066 (0.0031)
Sex	-0.3489 (0.1044)	0.2044 (0.3478)	-0.3673 (0.0731)	-0.0890 (0.1066)	-0.3495 (0.0516)
Black	0.1784 (0.1331)	0.8692 (0.5035)	-0.0177 (0.0910)	0.2487 (0.1551)	-0.0377 (0.0645)
Nonwhite/nonblack	1.4836 (0.2177)	3.1078 (0.4557)	0.3453 (0.1792)	1.4001 (0.1430)	0.2761 (0.1337)
Education	0.1426 (0.0929)	0.4393 (0.2928)	0.0522 (0.0681)	0.0937 (0.0895)	0.0108 (0.0512)
Marital Status	0.3153 (0.1317)	0.1223 (0.4279)	0.3049 (0.0933)	0.0984 (0.1332)	0.2509 (0.0695)
Urban	0.4474 (0.1451)	1.9717 (0.4662)	0.0212 (0.1066)	0.5250 (0.1502)	0.0451 (0.0721)
Midwest	-0.2501 (0.1160)	-0.9518 (0.3744)	-0.0454 (0.0810)	-0.2318 (0.1150)	-0.0021 (0.0612)
South	-0.0958 (0.1164)	-0.8377 (0.3687)	0.0642 (0.0824)	-0.1372 (0.1146)	0.0856 (0.0608)
West	0.1859 (0.1190)	0.2956 (0.3590)	0.1266 (0.0861)	0.1326 (0.1119)	0.1081 (0.0637)

Table 4. Continued

Variable	Tobit Model	Double-Hurdle Model		Purchase Infreq. Model	
	β_i	β_i	θ_i	β_i	θi
Season	-0.3724 (0.0770	-1.1025 (0.2517)	-0.1628 (0.0540)	-0.2898 (0.0788)	-0.1415 (0.0394)
Children < 5	-1.0018 (0.5374)	-4.2207 (1.9484)	-0.1862 (0.3923)	-0.9373 (0.5468)	-0.1096 (0.2996)
Children 5 to 13	-0.7699 (0.3219)	-1.5978 (0.9292)	-0.4352 (0.2295)	-0.5996 (0.2924)	-0.4155 (0.1708)
persons 14 to 24	-0.3635 (0.1752)	-1.3906 (0.6022)	-0.1703 (0.1134)	-0.1590 (0.1996)	-0.1608 (0.0787)
Persons 45 to 65	0.0895 (0.2016)	-0.6074 (0.6406)	0.2823 (0.1438)	-0.0353 (0.2049)	0.3757 (0.1054)
Persons > 65	-0.0689 (0.2860)	-2.1951 (0.9251)	0.4740 (0.2043)	-0.2467 (0.2907)	0.7394 (0.1529)
Variance	4.1119 (0.1245)	9.6261 (0.5141)	- -	2.7240 (0.0492)	2.7239 (0.0492)
Log likelihood	-5437.9	-5173.3		-6498.5	
IM statistic	272.2	139.5			

computer-software programing language was used to conduct the estimation. Since both the first and second derivatives of the log-likelihood function of the Tobit model are easily obtained, maximum likelihood estimates of the Tobit model were obtained via the method of Newton which uses the first and second derivatives of the log likelihood function. For the Double-Hurdle and Purchase infrequency model, however, the method of scoring (method of Berndt, Hall, Hall, and Hausman) which uses only the first derivatives was utilized. Least squares estimates were used as starting values for β, while estimates generated from a Probit among observations above and below the limit provided starting values for θ. Recall that in the Tobit model both the decision of whether to purchase and how much to purchase are captured in the β parameters, while in the Double-Hurdle model the decision of whether to purchase is embodied in θ, and β embodies the second decision of how much to purchase. With regard to the Purchase Infrequency model, θ is associated with the probability of infrequent purchases, while β reflects the decision of how much to purchase.

In Table 4, the estimated coefficient of each variable is presented. The estimated standard error for each coefficient is given in parentheses. Also present is the variance of the error term and the log-likelihood ratio associated with each model. In addition the information-matrix test statistic is computed for the Tobit and Double-Hurdle model.

With the exception of the sex and the household composition variable associated with the proportion of persons in the household between 45 to 65 years of age, the signs of the β coefficients are uniform across models. The β coefficient associated with the sex variable was negative in the Tobit and Purchase Infrequency model, but positive in the Double-Hurdle model. Past studies have indicated that females have a tendency to spend more on fresh vegetables than men, thus a negative coefficient would be more in line with the results of previous studies. In contrast to the Tobit model which indicated that persons 45 to 65 years old spend more on fresh-winter vegetables than those of 25 to 44 years, the sign of the corresponding β coefficient in the Double-Hurdle and Purchase Infrequency model implied the reverse. Among those variables with β coefficients whose signs were uniform across models, the coefficients associated with the household size and household size squared variables were the only ones with signs opposite to expectations. Household size was expected to have a positive impact on fresh vegetable expenditures, however the combined effect of the household-size and household-size-squared variables at the mean of the data is negative. The β coefficient of five variables are both at least twice as large as their corresponding standard errors, and have consistent signs across models. These include the food expenditure, nonblack/nonwhite, urban, Midwest region and season variable. In the Tobit model, 11 out of 19 β coefficients are at least twice the size of their standard errors, while the same holds true for 10 and 8 variables in the Double-Hurdle and Purchase Infrequency model, respectively. With regard to θ, the coefficient of 8 and 11 variables are at least twice the size of their standard errors in the Double-Hurdle and Purchase Infrequency model, respectively.

The log-likelihood ratio value provides a clue as to how well the models fit the data. But a direct comparison (based on the log-likelihood ratio) of the Purchase Infrequency model with either the Tobit or Double-Hurdle model is not possible because the models are not nested. However, the large amounts (over a 1000) by which the log likelihood ratio of both the Tobit and Double-Hurdle model exceed that of the Purchase Infrequency model (especially since the later include 20 more variables than the Tobit model) seems to indicate that the Tobit and Double-Hurdle model fit the data better than the Purchase Infrequency model. This result is not surprising, because the Purchase Infrequency model assumes that fresh vegetables are purchased Infrequently and hence it is likely that observe expenditures will deviate from actual consumption levels. However, since fresh vegetables are highly perishable it is unlikely that households store fresh vegetable items beyond a two-week period, implying frequent rather than infrequent purchases. Thus the results of the purchase infrequency model seem to validate the later explanation of households fresh vegetable purchasing behavior.

These results suggest that the Purchase Infrequency model may be more appropriate in the case of weekly as oppose to biweekly data. The results of the Purchase Infrequency model applied to weekly data is presented in Table 5. There is a substantial decrease in the size of the log-likelihood from that of Table 4, however, this change can be attributed mainly to an increase in number of observations (increase from 3368 to 6002 observations) that resulted when weekly data was used in lieu of biweekly data. The β coefficient associated with persons 45

Table 5. A Revisit of the Purchase Infrequency Model.

Variable	Purchase Infrequency Model	
	β_i	θ_i
Constant	-2.1492 (0.3018)	1.0676 (0.0925)
(Food Exp.)$^{1/2}$	0.6241 (0.0138	0.2743 (0.0060)
Household Size	-0.2728 (0.1138)	0.1342 (0.0397)
(Household Size)^2	0.0291 (0.0114)	-0.0210 (0.0041)
Age	0.0052 (0.0055)	-0.0063 (0.0020)
Sex	-0.1088 (0.1057)	-0.2790 (0.0365)
Black	0.2931 (0.1473)	-0.0206 (0.0476)
Nonwhite/nonblack	1.2270 (0.1780)	0.0340 (0.0727)
Education	0.1281 (0.0901)	-0.0105 (0.0331)
Marital Status	0.0346 (0.1281)	0.1154 (0.0461)
Urban	0.5190 (0.1596)	0.0458 (0.0489)
Midwest	-0.3121 (0.1174)	0.0643 (0.0399)
South	-0.2305 (0.1148)	0.0472 (0.0398)
West	0.0738 (0.1157)	0.1769 (0.0414)

Table 5. Continued

Variable	Purchase Infrequency Model	
	β_i	θ_i
Season	-0.3168 (0.0799)	-0.0940 (0.0265)
Children < 5	-1.4157 (0.5474)	-0.3465 (0.1940)
Children 5 to 13	-0.6503 (0.3107)	-0.3359 (0.1096)
Persons 14 to 24	-0.1669 (0.1972)	-0.2039 (0.0577)
Persons 45 to 65	0.0882 (0.2047)	0.3782 (0.0696)
Persons > 65	-0.0577 (0.2798)	0.6841 (0.0993)
Variance	3.7162 (0.0605)	
Log Likelihood	10904.9	

to 65 years old changed from negative to positive, also the β coefficient associated with black household heads, households located in the South, and children less than 5 years old, switched to being at least twice the size of their corresponding standard errors. With regard to the θ coefficients, the coefficients associated with the intercept and households located in the Midwest switch from negative to positive. The coefficients associated with household size, households located in the West and children less than 5 years old have changed to being at least twice the size of their corresponding standard errors, while the opposite has occurred in the case of the variables represented by households headed by nonwhite/nonblacks and those headed by high school graduates.

Concluding that the purchase infrequency model is inconsistent with the data generating process, leads to a comparison between the Tobit and Double-Hurdle model. In this case, however, a formal test based on the log likelihood ratio's of the two models can be constructed to test the Tobit specification against the Double-Hurdle model, because the Tobit model is a special case of the Double-Hurdle model. Specifically, the Double-Hurdle model is reduced to the Tobit model when $\theta=\beta/\sigma$, thus the nested test involving the two models is a test of the null hypothesis that $\theta=\beta/\sigma$. To test this hypothesis, the likelihood ratio test statistic which is distributed asymptotically as chi-square with 20 degrees of freedom was calculated as 529.2. Comparing this computed value with the critical chi-square statistic value at the .01 level, leads to a rejection of the null hypothesis that the restrictions embodied in the Tobit model are valid. However, such a conclusion is acceptable only if the alternative model (the Double Hurdle model) is correctly specified.

Subsequently, the information matrix (IM) misspecification test based on the unique elements of the information matrix was performed on both the Tobit and Double-Hurdle model. The elements chosen for the test corresponded to the variance of the error term (σ^2), variances of the β's, and the covariances between σ^2 and the β's. Like the LR test statistic, the IM test statistic is also distributed asymptotically as a chi-square. The IM statistic with 41 degrees of freedom for the Tobit and Double-Hurdle model was computed at 272.2 and 139.52, respectively. Given that the critical value at the .01 level of a chi-square statistic with 41 degrees of freedom is 64.95, the null hypothesis of correct model specification was rejected in both models.

As pointed out in chapter 3, a possible source of misspecification in the Tobit and Double-Hurdle model is non-normally distributed disturbance terms. Given this possibility, the inverse-hyperbolic-sine transformation was applied to both models. In imposing the transformation, the constant term in the previous specification was dropped because α (the location parameter) is not identified in the presence of an exogenous variable with zero variance (Ramirez et al., 1988).

The results of the models (Tobit and Double-Hurdle model) with the IHS transformation are presented in Table 6. The estimate of α is significantly different from zero in both models, implying that the dependent variable enters the models nonlinearly. The IHS transformation has introduced several changes in the estimated β coefficients in both

Table 6. Modifications of the Tobit and Double-Hurdle Model.

Variable	IHS-Tobit Model	IHS-Double-Hurdle Model		IHS-Tobit with Heteroskedasticity
	β_i	β_i	θ_i	β_i
Constant (or α)*	0.4365* (0.0244)	0.4327* (0.0485)	1.4435 (0.2032)	0.3860* (0.0240)
(Food exp.)$^{1/2}$	0.4330 (0.0152)	0.3565 (0.0401)	0.3873 (0.0136)	0.4430 (0.0150)
Household size	0.0311 (0.0675)	0.1952 (0.1004)	0.0227 (0.0936)	0.0390 (0.0710)
(Household size)^2	-0.0089 (0.0071)	-0.0199 (0.0101)	-0.0106 (0.0114)	-0.0100 (0.0080)
Age	0.0101 (0.0029)	0.0274 (0.0048)	-0.0020 (0.0041)	0.0110 (0.0030)
Sex	-0.1996 (0.0607)	0.1603 (0.0963)	-0.3673 (0.0732)	-0.2230 (0.0610)
Black	0.1081 (0.0784)	0.2790 (0.1306)	-0.0177 (0.0910)	0.1080 (0.0770)
Nonwhite/nonblack	0.7604 (0.1324)	0.8767 (0.1820)	0.3453 (0.1793)	0.7780 (0.1400)
Education	0.1640 (0.0521)	0.3341 (0.0829)	0.0522 (0.0681)	0.1690 (0.0520)
Marital Status	0.1744 (0.0779)	-0.0996 (0.1163)	0.3049 (0.0933)	0.1720 (0.0800)
Urban	0.3764 (0.0819)	0.9520 (0.1437)	0.0212 (0.1068)	0.3990 (0.0830)
Midwest	-0.1749 (0.0691)	-0.3438 (0.1128)	-0.0454 (0.0810)	-0.1830 (0.0920)
South	-0.0861 (0.0693)	-0.2627 (0.1098)	0.0642 (0.0825)	-0.0920 (0.0700)
West	0.1255 (0.0709)	0.1060 (0.1054)	0.1267 (0.0862)	0.1140 (0.0720)

Table 6. Continued

Variable	IHS-Tobit Model	IHS-Double-Hurdle Model		IHS-Tobit with Heteroskedasticity
	β_i	β_i	θ_i	β_i
Season	-0.2396 (0.0458)	-0.2303 (0.0744)	-0.1628 (0.0540)	-0.2310 (0.0460)
Children < 5	-0.4772 (0.3171)	-0.7596 (0.5326)	-0.1862 (0.3925)	-0.4860 (0.3310)
Children 5 to 13	-0.3406 (0.1901)	-0.1825 (0.2852)	-0.4352 (0.2296)	-0.3250 (0.1990)
Persons 14 to 24	-0.1879 (0.1004)	-0.1625 (0.1698)	-0.1703 (0.1136)	-0.1420 (0.1020)
Persons 45 to 65	0.0871 (0.1187)	-0.1723 (0.1812)	0.2823 (0.1439)	0.0840 (0.1220)
Persons > 65	-0.0303 (0.1664)	-0.7343 (0.2659)	0.4740 (0.2045)	0.0130 (0.1690)
Variance Parameters				
0	1.4823 (0.0811)	1.5772 (0.2388)		0.4430 (0.1330)
1	-			0.1110 (0.0240)
2	-			0.5390 (0.1200)
Log Likelihood	-5103.1	-5096.3		-5078.9
IM Statistic	94.17	97.70		50.89

the Tobit and Double-Hurdle Model. The size of the β coefficients in the IHS models are considerably smaller than in the previous specification. The coefficients associated with household size and Household size squared have switch signs in accordance with priori expectations. But in the case of the Tobit model the coefficients of these variables have switch to being less than twice the size of their corresponding standard errors. The coefficient of the education variable have changed to being twice the size of its standard error in both models. Although the standard error associated with the marital status variable remains large, the sign of its β coefficient in the Double-Hurdle model has changed from positive to negative, which is contrary to expectations. Among the household composition categories, the coefficient of the variable corresponding to the proportion of persons 14 to 24 years old is now less than twice the size of its standard error in both models, while that associated with the proportion of household members less than 5 years old in the Double-Hurdle model have changed in similar fashion.

The transformation also brought about an increase in the log-likelihood of both models. However, the increase associated with the Tobit model (334.8) was much greater than that of the Double-Hurdle (77) model. The LR test statistic for testing the IHS-Tobit model against the IHS-Double-Hurdle model was estimated at 13.6 with 20 degrees of freedom. A comparison with the corresponding critical value at the 0.1 level fails to reject the IHS-Tobit specification. This contrast with the untransformed models which gave rise to the opposite conclusion of rejecting the Tobit specification. These results seem to indicate that non-normality was a much more serious problem in the Tobit model compared with the Double-Hurdle model. The information matrix test was again applied to the transformed models. The IHS-Tobit and Double-Hurdle models yielded values of 94.17 and 97.70, respectively, with 39 degrees of freedom (two degrees of freedom were lost when the intercept was dropped). With a critical value of 63.69 at the .01 level both models were again deemed misspecified. Note however, that the computed IM statistic represents a substantial decrease, especially in the case of the Tobit model, from the previous IM estimate. Thus the IHS transformation has corrected for some of the misspecification.

However, since the IHS transformation did not completely correct for the misspecification, sources of misspecification other than non-normality were considered. A second likely source of misspecification is heteroscedastic disturbance terms. Some experimentation suggested that indeed the variance was not constant over the households in the sample. Although several other regressors were analyzed, attention centered on modelling the variance as a function of food expenditure and household size and/or composition. The IHS-Tobit model was considered first for the incorporation of the heteroscedastic disturbance structure. The heteroscedastic specification which produced the greatest improvement in the log likelihood of the Tobit model is presented in column 5 of Table 6. The variance was modeled as

$$\sigma_i^2 = \tau_0 + \tau_1 Z_{1i} + \tau_2 Z_{2i} \tag{56}$$

where Z_1 was defined as the square root of household food expenditures and Z_2 was the proportion of the household that was between 14 and 65 years old. The heteroscedastic structure had very little impact on the signs, magnitude and level of significance of the estimated coefficients from the IHS-Tobit model. However, the LR statistic to test the hypothesis that $\tau_1 = \tau_2 = 0$ has a value of 48.4 with 2 degrees of freedom. A comparison with a critical value with 2 degrees of freedom at any probability level will result in a rejection of the null hypothesis, thus implying that accounting for heteroscedasticity did significantly improve the fit of the Tobit model. Moreover, the information matrix test produced a value of 50.89 with 39 degrees of freedom. Upon comparing this computed value with a tabled value of 54.57 at the .05 level, the hypothesis of proper specification was not rejected.

The foregoing discussion seems to suggest that a properly specified Tobit model provides an appropriate representation of households fresh-winter vegetable consumption behavior. Implicitly, the results indicate that when non-normality and heteroscedasticity were accounted for, the Tobit censoring rule (zero expenditures on fresh winter vegetables results from corner solutions) adequately explains the realization of zero expenditures. As a consequence, the results of the IHS-heteroscedastic-Tobit model given in Table 6 were used to analyze the impact of economic and demographic variables on fresh-winter vegetable expenditures. Also, the results were used to forecast fresh-winter vegetable expenditures.

<u>Interpretation of the IHS-Heteroscedastic-Tobit Model</u>

In interpreting the results of the IHS-heteroscedastic-Tobit model, it is important to keep in mind that, because the model was selected based on priori misspecification testing, the sampling distribution of the associated estimator is unknown. Consequently, the reported standard errors may be misleading and standard hypothesis testing is not appropriate.

The IHS-Heteroscedastic Tobit model suggested qualitative impact of the explanatory variables on fresh-winter vegetable expenditures is as expected and for the most part is consistent with the findings of previous studies. The coefficient of both food expenditures (household income) and the age of the household head are at least twice the size of their corresponding standard errors. The positive sign associated with the coefficient on household size suggest that larger households spend more on fresh vegetables, however, according to the negative sign on the coefficient of household size squared, there are economies of scale in consumption, since increases in expenditures resulting from household size increases at a decreasing rate. The coefficients associated with the following variables; sex, race, educational level and marital status of the household head, were all at least twice as large as their corresponding standard errors. Households headed by females who are nonwhite/nonblack, married, and are high school graduates, tend to spend significantly more on fresh-winter vegetables than others. Location is also an important determinant of fresh-winter vegetable expenditures. According to the results, Urban dwellers spend a significantly greater amount on fresh vegetables than do their rural counterpart. The greater incidence of home gardens in rural areas is probably a partial explanation for this result. With regard to the regional location of the household, the results of the IHS-heteroscedastic-Tobit model suggest that households located in the West spend more on fresh vegetables than

Northeastern households, but the Northeast spend more than either the South or the Midwest, while the Midwest has the least tendency to spend on fresh-winter vegetables. Finally, the results depict a definite pattern between the age composition of the household and its expenditures on fresh-winter vegetables. Household expenditures increases continuously along with the age of household members until it reaches a peak that corresponds with the 45 to 65 age group.

Incorporating the heteroscedastic structure in the Tobit model left the estimated coefficients largely unchanged, but the inverse-hyperbolic transformation did have a considerable impact on the magnitude of these coefficients. Table 7 illustrates how the transformation acts on the extreme values of the dependent variable. For values of y near the mean over the entire sample or over the sub sample associated with positive fresh-winter vegetable expenditures, the transformation has little effect. However the sample contained 22 observations with y exceeding five times the mean of non-limit households and six observations with y exceeding ten times the mean of non-limit households. For these observations the transformation acts to reduce their magnitude relative to those around the mean. Consequently, the influence of these observations on the coefficient estimates were reduced.

With the use of equations 44 and 48 of chapter 3, the results of the IHS-heteroscedastic-Tobit model can be used to predict the impact of the explanatory variables on household expenditure levels. These predicted effects are presented in the second column of Table 8. The values for the continuous variables (food expenditure, household size

Table 7. The Effect of the IHS Transformation on the Dependent Variable

	Value of Y	I(Y): $\alpha = 0.386$
Mean of Y	1.513	1.438
Mean of Y > 0	2.235	2.023
5 * Mean of Y > 0	11.180	5.618
10 * Mean of Y > 0	22.350	7.387

Table 8. The IHS-Heteroscedastic-Tobit Model Suggested Impact of Socioeconomic Variables on Fresh-Winter Vegetable Consumption

Variable	Simulated Impact	Percentage Change From Base		Elasticities
		base	%Δ	
Food Exp.	0.0752	-		1.96
Household Size	-0.0280	-		-0.05
Age	0.0235	-		0.72
Sex	-0.4821	5.2185	-9.24	-
Race				
Black	0.2318	4.8148	4.81	-
Nonwhite/nonblack	1.8837	4.8148	39.12	-
Education	0.3657	4.6343	7.89	-
Marital Status	0.3653	4.6841	7.80	-
Urban	0.8072	4.1765	19.33	-
Region				
Midwest	-0.3836	4.9785	-7.71	-
South	-0.1958	4.9785	-3.93	-
West	0.2515	4.9785	5.05	-
Season	-0.4908	5.1165	-9.59	-
Household Composition				
Children < 5	-0.9690	4.9827	-19.45	-
Children 5 to 13	-0.6656	4.9827	-13.36	-
Persons 14 to 24	-0.2999	4.9827	-6.02	-
Persons 45 to 65	0.1845	4.9827	3.70	-
Persons > 65	0.0282	4.9827	0.57	-

and age) were obtained by evaluating equation 48 at the means of the data.

The predicted effect associated with a discrete variable (the Southern region variable, for example), on the other hand, was estimated as the difference in value between equation 44 evaluated with the value of the variable in question (South) set equal to one, while assigning a zero value to the other discrete variables in the group (the West and Midwest are assign zero values) and keeping all other variables at their means, and equation 44 evaluated with all variables (the South, Midwest and West) in the group set equal to zero, while all other variables are kept at their means. So, in effect, with regard to the discrete variables, column 2 expresses the change in expenditures that result when the variable in question (South) is other than the omitted base variable (Northeast). Column 4, in turn, expresses the changes associated with individual discrete variables as a percentage of the expenditure level associated with the base variable. Finally, column 5 provides expenditure elasticities for the continuous variables.

Food expenditure elasticity was estimated at 1.9, implying that a 10 percent increase in home food expenditures is accompanied by an estimated 19 percent increase in fresh-winter vegetable expenditures. This result is in sharp contrast with previous studies which found fresh vegetables to be income inelastic. Keep in mind, however, that in this study food expenditures is used in place of income, and while previous studies grouped winter and other fresh vegetables in one category, this study focuses mainly on vegetables consumed during the winter months. A negative elasticity was estimated for household size. According to the estimated elasticity associated with the age of the household head, a 10 percent increase in age results in a 7.2 percent increase in expenditures on fresh-winter vegetables.

With regard to the discrete variables, households headed by males spend an estimated 9.2 percent less than their female counterparts, blacks and nonblacks/nonwhites, respectively, spend an estimated 4.8 and 39.1 percent more than whites, high school graduates spend an estimated 7.9 percent more than non-high school graduates, while the expenditures of households with married couples are an estimated 7.8 percent above those of households with no married couples. Similarly, urban households' fresh-winter vegetable expenditures are an estimated 19.3 percent in excess of the expenditures of rural households. Differences in expenditures also exist between regions. For example, while households located in the Midwest and the South spend an estimated 7.7 and 3.9 percent less, respectively, than Northeastern households, households residing in the West spend an estimated 5.1 percent more than those in the Northeast. The estimated 9.6 percent difference between the expenditures of household surveyed in the months of November and December and that of households surveyed during the remaining months (March, April, May, June), suggests that seasons do influence fresh-winter vegetable expenditures. Differences across seasons with regard to the availability of fresh vegetables may help explain the difference in expenditure levels. Finally, the estimates indicate that compared with persons 25 to 44 years old, fresh-winter vegetable expenditures is 19.5 percent less for those less than 5 years, 13.4 less for those between 5 and 13, 6.0 percent less for those between 14 and 24, 3.7 and 0.6 percent more, respectively, for those 45 to 65 years and those over 65.

Results of the Heckman Two-Step Estimation Procedure

Heckman two-step estimates of the Tobit model are given in Table 9. The results when Ordinary Least Squares (OLS) is used in the second stage of the estimation is given in column 2. Alternatively, column 3 contains the estimates of the Tobit model, when Generalized Least Squares (GLS) as oppose to OLS is used in the second stage. The inverse of the square root of the result of equation 53 was the weight used in conducting the GLS estimation. As a basis for further comparison with the results of chapter 4, the IHS transformation was employed along with the GLS estimator. These results are shown in column 4. Because of the transformation, however, the results of column 4 are not directly comparable with that of the previous columns. The coefficients reported in column 4 are interpreted as $\partial E[I(y^*)]/\partial x_j$. But what is needed is $\partial E(y^*)/\partial x_j$. Given that $E[(y^*)]=x\beta=f(x)$, we define $F=I(y)-f(x)=0$. Thus,

$$\frac{\partial F}{\partial y}\frac{dy}{dx} + \frac{\partial F}{\partial x}\frac{dx}{dx} = 0, \quad \text{consequently} \quad \frac{dy}{dx} = \frac{-\ \partial F/\partial x}{\partial F/\partial y}$$

From the above expressions it follows that $\partial E(y^*)/\partial x_j = \beta(\alpha^2 y^2 + 1)^{1/2}$. Thus, evaluating the dependent variable y at its mean, $(\alpha^2 y^2 + 1)^{1/2} = 1.283$, is the factor by which the estimated parameters have to be adjusted to account for the IHS transformation. The results of this adjustment is given in column 5.

Table 9. Heckman Two-Step Estimator

Variable	OLS	GLS	GLS/IHS	GLS/IHS*
Constant	-5.0717 (0.5357)	-4.6695 (0.4980)	- -	- -
$(\text{Food exp.})^{1/2}$	0.8994 (0.0580)	0.8623 (0.0530)	0.6852 (0.0449)	0.8783
Household size	-0.2900 (0.1385)	-0.2431 (0.1251)	-0.1657 (0.0735)	-0.2124
$(\text{Household size})^2$	0.0233 (0.0170)	0.0163 (0.0148)	0.0140 (0.0086)	0.0179
Age	0.0110 (0.0056)	0.0111 (0.0049)	0.0078 (0.0030)	0.0010
Sex	-0.3976 (0.1203)	-0.4678 (0.1049	-0.1770 (0.0688)	-0.2269
Black	0.2402 (0.1241)	0.1538 (0.1037)	0.2223 (0.0773)	0.2850
Nonwhite/nonblack	1.7032 (0.3177)	1.5004 (0.2666)	1.4118 (0.1943)	1.8097
Education	0.1858 (0.0891)	0.1700 (0.0738)	0.1464 (0.0539)	0.1877
Marital Status	0.3954 (0.1370)	0.4296 (0.1216)	0.0945 (0.0745)	0.1211
Urban	0.5382 (0.1379)	0.4600 (0.1295)	0.5708 (0.0868)	0.7317
Midwest	-0.3159 (0.1106)	-0.3285 (0.0922)	-0.3693 (0.0714)	-0.4734
South	-0.1366 (0.1136)	-0.1476 (0.0953)	-0.2573 (0.0709)	-0.3298
West	0.2329 (0.1199)	0.2144 (0.1033)	0.1571 (0.0772)	0.2014
Season	-0.4979 (0.0818)	-0.4297 (0.0701)	-0.3968 (0.0552)	-0.5086

Table 9. Continued

Variable	OLS	GLS	GLS/IHS	GLS/IHS*
Children < 5	-1.2776 (0.4798)	-1.1491 (0.3822)	-1.4916 (0.3137)	-1.9120
Children 5 to 13	-0.9705 (0.3660)	-0.9076 (0.2928)	-0.8421 (0.2197)	-1.0795
Persons 14 to 24	-0.5290 (0.1912)	-0.4763 (0.1582)	-0.4660 (0.1250)	-0.5974
Persons 45 to 65	0.1161 (0.2055)	0.0937 (0.1864)	0.0550 (0.1267)	0.0705
Persons > 65	-0.0529 (0.2891)	-0.0963 (0.2641)	-0.2321 (0.1716)	-0.2975
λ	3.1015 (0.3444)	2.7865 (0.2931)	2.0882 (0.2507)	2.6768

The sign of the estimated coefficient of each variable is the same for both the OLS Heckman two-step estimator (subsequently referred to as the OLS estimator) and the Maximum Likelihood estimator of the Tobit model (Table 4). However, with few exceptions, the estimates of the OLS estimator are of larger magnitude. Also, contrary to the Maximum Likelihood estimator, the standard errors associated with the coefficients of the following variables; if household head is black, if household head is a high school graduate, if household is located in the west, and children less than 5 years old, are at least twice the size of their corresponding coefficients in the OLS estimator.

As indicated above, the OLS Heckman two-step estimator is inherently heteroscedastic. Thus the GLS Heckman two-step estimator (subsequently referred to as the GLS estimator) is more efficient. The results of the two estimators are quite comparable, however. Corresponding coefficients are of the same sign. In fact, the signs of the coefficients are uniform across the three specifications presented in Table 9. Furthermore, most of the variables enter the two specifications at similar levels of significance. However, as was expected, the estimated coefficients of the GLS estimator were in general of smaller magnitude than that of the OLS estimator.

A comparison of the GLS estimator with the GLS/IHS (GLS estimator with the inverse hyperbolic transformation) estimator revealed that the IHS transformation has brought about a change in the significance level of several variables. For example, the coefficients of the following variables; household size squared, if household head is black and if household resides in the South, switch to being at least twice the size of their corresponding standard errors, while the coefficient of the marital status variable switch to being less than twice the size of its standard error. Recall that when the IHS transformation was employed, using the Maximum Likelihood estimator, the sign of the coefficient associated with household size and the proportion of household members less than 65 years old, switched from negative to positive, while the sign of the household size squared coefficient switched from positive to negative. In the case of the Heckman two-step estimator, however, the signs of the coefficients remained unchanged.

The coefficient associated with λ (the inverse of mills ratio) provides an indication of whether deleting the observations corresponding to zero expenditure levels results in biased parameter estimates (selectivity bias). The coefficient is significant across the specification presented in Table 9, implying that if the observations associated with zero expenditures are ignored in the estimation process bias parameter estimates will result.

Fresh-Winter-Vegetable Projections

This section provides projected percentage changes in fresh winter vegetables resulting from increases in food expenditures and changes in the proportion of the population by marital status, race, region, and age. The projected changes in these independent variables are shown in Table 10. The projections for food expenditures were based on the assumption that food expenditures would increase by 2 percent per year in real terms. Population projections by marital status, race, region and age were adopted from the middle series projections provided by the Bureau of Census. Table 10 displays definite patterns of demographic

Table 10. Projected Home Food Expenditure, Number of Households, proportion of Households with Married Couples, and Proportion of the Population by, Age, Race and Region.

Variables	Years				
	1985	1990	1995	2000	2010
Food Expenditure	20.83	23.02	25.42	28.07	34.21
Household # (mill.)	86789	94227	100308	105933	117526
Marital Status	0.580	0.563	0.547	0.531	0.499
Race					
White	0.851	0.844	0.838	0.831	0.817
Black	0.122	0.126	0.130	0.133	0.141
Nonwhite/nonblack	0.027	0.030	0.032	0.036	0.042
Region					
Northeast	0.209	0.202	0.198	0.194	0.186
Mid-west	0.248	0.239	0.231	0.223	0.209
South	0.343	0.349	0.356	0.362	0.372
West	0.200	0.209	0.216	0.222	0.233
Age					
Less than 5	0.077	0.077	0.072	0.066	0.063
5 to 13	0.124	0.129	0.133	0.128	0.113
14 to 24	0.182	0.155	0.145	0.149	0.151
25 to 44	0.309	0.326	0.318	0.299	0.263
45 to 65	0.187	0.186	0.202	0.227	0.275
over 65	0.120	0.127	0.131	0.130	0.138

Source: U.S. Department of Commerce, Bureau of the Census (various issues).

shifts. The proportion of households with married couples is projected to decline through out the projection period (1985 - 2010). Non-white races share of the population is on the increase while that of whites is decreasing. Regional shifts are also taking place. In terms of proportions, the population is shifting from the Northeast and Midwest to the South and the West. The age composition of the population, on the other hand, is shifting from ages less than 45 to ages above 45. Thus given the qualitative results of the IHS-heteroscedastic-Tobit model we can expect the projected changes in food expenditures, racial, regional and age composition population shifts to all have a positive impact on future fresh-winter vegetable expenditures. In contrast, the projected changes in the marital status of households can be expected to have a negative impact on projected expenditures.

The projected impact of these variables and increases in number of households (or increases in population give household size) on fresh-winter vegetable consumption are shown in Table 11. Equation 44 of chapter 3 was used for the simulation. The projections are based on two main assumptions. First, the analysis assumes that the relationship that exist between fresh-winter vegetable expenditures and food expenditures along with the demographic variables remains unchanged over time. Second, as consumers economic and demographic circumstances change, it is assumed that they acquire the consumption behavior of individuals already observed in the new circumstance.

The projections are all in line with expectations. As a result of growth in food expenditures, household-fresh-winter vegetable expenditures in the year 2010 can be expected to be 30.3 percent above

Table 11. Projected Effect on Fresh-Winter Vegetable Household Expenditures due to Changes in Food Expenditures, Household Number, Proportion of Households with Married Couples, and Changes in Proportion of the Population by, Race Region, and age.

Variables	Years				
	1985	1990	1995	2000	2010
	-------- percentage of 1985 value -----------				
Food Expenditure	100	105.1	110.6	116.6	130.3
Marital Status	100	99.9	99.7	99.6	99.3
Race	100	100.1	100.2	100.4	100.7
Region	100	100.1	100.2	100.3	100.4
Age	100	100.1	100.3	100.6	101.1
Combined Effect	100	105.3	111.1	117.6	132.1
Household #	100	108.6	115.6	122.1	135.4
Total Effect	100	114.3	128.4	143.5	178.9

expenditures in 1985. Changes in the marital status of households is projected to cause fresh winter vegetable expenditures to decrease by 0.7 percent between 1985 and 2010. Racial population shifts can be expected to bring about a 0.7 percent increase in expenditures over the same period. Similarly, Regional and age composition population shifts are anticipated to result, respectively, in a 0.4 and 1.1 percent increase in fresh-winter vegetable expenditures. The combined effect of changes in these economic and demographic variables is to cause expenditures to increase by 32.1 percent over the projection period. Comparing the combined effect with the separate effects, it is apparent that changes in food expenditures account for most of the increase in expenditures. This result provides justification for considering only prices and income when analyzing time series data. Demographic factors changes so slowly over time that they are generally assumed constant. However, as the results shown in Table 8 point out, consumption patterns do differ significantly across demographic groups.

Increases in number of households (population growth) had an even greater impact on consumption than changes in food expenditures. Changes in number of households is predicted to cause fresh winter vegetable expenditures to increase by 35.4 percent between 1985 and 2010. When population growth projections are combined with the other projections (food expenditure, marital status, race, region and age), fresh winter vegetable consumption is expected to increase by 78.9 percent over the projection period.

CHAPTER 5
SUMMARY AND IMPLICATIONS

The primary objective of this study was to analyze the impact of socioeconomic and demographic factors on the consumption of fresh-winter vegetables. Data generated from the 1984 diary survey of the Continuing Consumer Expenditure Survey, sponsored by the Bureau of Labor Statistics, were used for the study. The data were comprised of individual household's expenditures on fresh vegetables along with information characterizing the household's economic and demographic situation. A significant portion (one-third to be exact) of the households in the sample of interest reported zero expenditures on fresh-winter vegetables. This phenomenon of a large proportion of observations on the dependent variable taking on zero values renders standard regression methods an inappropriate empirical framework. Depending on the explanation given for households not purchasing the item during the survey period, several censored regression models have been developed to account for the occurrence of zero expenditures.

The Tobit model assumes that the household did not purchase the good during the survey period because the household did not desire or do not consume the good. Furthermore the model allows the same set of parameters and variables to characterize the decision of whether or not to purchase and the decision of how much to purchase. The Double-Hurdle model, on the other hand, recognizes, that it is possible for the household to desire the good but impediments to consumption may prohibit purchases. Thus the Double-Hurdle model conceptualizes the household's purchasing behavior into two steps or what Cragg calls hurdles. First the consumer decides whether or not to purchase, and second decides on how much to purchase. Consequently the model allows a different set of parameters to characterize each decision. And, if it turns out that the Tobit censoring rule is the correct interpretation of the household's consumption behavior, then the Double-Hurdle model reduces to the Tobit model. The Purchase Infrequency model provides yet another explanation for observing zero expenditures. The model assumes that the households always consume the good in question, but because the good is purchased infrequently households reporting zero expenditures purchased the good before or after the survey period as oppose to within the survey period. Like the Double-Hurdle model, the Purchase Infrequency model allows one set of parameters to capture the purchase infrequency phenomenon and a different set of parameters to characterize the decision on how much to purchase. But unlike the Double-Hurdle model the Purchase Infrequency model is not a generalization of the Tobit model.

Specifying fresh-winter vegetable expenditures as a function of the square root of home food expenditures, household size, household size squared, the age, sex, race and marital status of the household head, whether or not the household head completed high school, the location of the household with regard to urban/rural and region of residence, the months of the year during which the household was surveyed and finally the age composition of the household, maximum likelihood estimates of all three models were obtained via the method of Newton in the case of the Tobit model and the modified scoring method in the case of the Double-Hurdle and Purchase Infrequency model. The log-likelihood of the Tobit (-5438) and Double-Hurdle (-5173) model far exceeded that of the Purchase Infrequency model (-6499). Although a casual comparison of log-likelihoods in that manner does not constitute a conclusive test, given that the Purchase Infrequency model assumes that the good is purchased infrequently, when in the case of fresh vegetables frequent rather than infrequent purchases seems to be the case, the disparity that exist in the log-likelihoods (especially since the Tobit model has 20 less variables as the Purchase Infrequency model), was construed as evidence in support of priori suspicion that the Purchase Infrequency model is inconsistent with households fresh-winter vegetable consumption behavior.

Next, the log-likelihood ratio test statistic was constructed to test the Tobit specification against the Double-Hurdle model. The test led to a rejection of the null hypothesis that the restrictions embodied in the Tobit model are valid. However, upon application of the Information Matrix (IM) misspecification test to the Double-Hurdle and the Tobit model, both models were deemed misspecified.

In considering sources of misspecification, non-normality was the first to be visited. Assuming that the disturbance terms are probably non-normally distributed, the inverse-hyperbolic-sine (IHS) transformation, considered a transformation to normality, was applied to both models. The location parameter associated with the IHS transformation was significantly different from zero in both models, implying that the dependent variable enters the models nonlinearly. Furthermore, the transformation brought about a substantial increase in the log-likelihood of the Tobit model relative to the increase the transformation gave rise to in the Double-Hurdle model. The improvement in the fit of the Tobit model was such that the LR test statistic for testing the IHS-Tobit specification against the IHS-Double-Hurdle model failed to reject the Tobit specification. However, despite this improvement the IM test (once again) indicated that the IHS-Tobit along with the IHS-Double-Hurdle model was misspecified.

This led to considering heteroscedasticity as a remaining source of misspecification. In fact some experimentation suggested that the variance was not constant over the households in the sample. To accommodate heteroscedasticity, the variance of the error term was modelled as a function of a constant, the square root of household food expenditures, and the proportion of the household that was between 14 an 65 years old. Considering that the LR test failed to reject the IHS-Tobit specification against the IHS-Double-Hurdle model, and realizing that the Tobit model presents less difficulty in incorporating the heteroscedastic disturbance structure, the Tobit model was considered first. Based on the LR test statistic the null hypothesis that the estimated parameters in the variance of the disturbance, associated with the square root of food expenditure and the proportion of the household between 14 and 65, are equal to zero was rejected, indicating that accounting for heteroscedasticity did improve the fit of the model. Furthermore, the information matrix test to test the null hypothesis of no misspecification in the IHS-Heteroscedastic-Tobit model failed to reject the null Hypothesis.

Concluding that the IHS-heteroscedastic-Tobit model was an appropriate representation of household's fresh-winter vegetable consumption behavior, the model was used to continue the analysis of the impact of demographic variables on fresh-winter vegetable consumption. The results of the IHS-heteroscedastic-Tobit model indicated that food expenditure (household income) had considerable impact on fresh-winter vegetable expenditures. A suggested 10 percent increase in food expenditures would result in an estimated 19 percent increase in fresh-winter vegetable expenditures. This result contrast with previous studies in that they estimated an inelastic income elasticity for fresh vegetables. Household size was not an important factor in explaining fresh-winter vegetable expenditures, and in contrast to most previous studies, the household size elasticity was negative. The age, sex, and marital status of the household head all had considerable impact on fresh-winter vegetable consumption. A 10 percent increase in the age of the household head would cause household expenditures to increase by an estimated 7.2 percent. Female headed households spend an estimated 9.2 percent more on fresh-winter vegetables than male headed households, while if the household head is at least a high school graduate the household would spend an estimated 7.9 percent more than a household whose head did not complete high school. Race was also an important determinant of vegetable expenditures. As a group races other than white and blacks spend an estimated 39.1 percent more on fresh vegetables than whites. In comparison, Black households spend an estimated 4.8 percent more than whites. Urban dwellers spend an estimated 19.3 percent more on fresh-winter vegetables than their rural counterparts. Region as another location variable also appears to affect household's expenditure levels. For example, while households in the Midwest spend an estimated 7.7 percent less than households residing in the Northeast, those in the South spend an estimated 3.9 percent less, and those in the Midwest an estimated 5.1 percent more. With regard to household age composition, expenditures seem to vary in direct proportion with the age of household members. Fresh-winter vegetable expenditures were estimated to be lowest for persons less than 5 years old (19.5 percent less than persons 25 to 44 years) and highest (3.7 percent more than persons 25 to 44 years) for persons between 45 and 65 years.

For the sake of comparison the Tobit model was also estimated with Heckman two step estimation procedure. In general, the results generated by that procedure were comparable with that of the maximum likelihood method, both in terms of magnitude and level of significance.

Finally, the results of the IHS-heteroscedastic-Tobit model was used to project percentage changes in fresh-winter vegetable expenditures from a 1985 base year to the year 2010. The projections were based on the assumption that at home food expenditures would increase by 2 percent per year in real terms. In addition to food expenditures, the fresh-winter vegetable projections were also conditioned on population growth and population projections by marital status, race, region and age composition. These population projections were obtained from the middle series projections provided by the Bureau of the Census.

The projections suggest that changes in food expenditures would cause fresh-winter vegetable expenditures by households to increase an estimated 30.3 percent between 1985 and 2010, changes in the proportion of households with married couples would result in an estimated decrease of 0.7 percent, changes in racial mix an estimated increase of 0.7 percent, regional shifts an estimated increase of 0.4 percent, and changes in population age composition an estimated increase of 1.1 percent. The estimated combined effect of all these changes on fresh-winter vegetable consumption is a rise in expenditure levels of 32.1 percent from the year 1985 to the year 2010. This projection, however, does not include population growth effects. In fact, in isolation, population growth was expected to cause fresh-winter vegetable expenditures to increase by an estimated 35.4 percent over the projection period. When population growth is combined with the other effects, expenditures on fresh-winter vegetables were projected to increase by an estimated 78.9 percent from 1985 to 2010.

This study suggests that misspecification in the Tobit model in the form of non-normality and heteroscedasticity, can lead to wrongly rejecting the Tobit model when testing against its generalization--the Double-Hurdle model. Furthermore, a correctly specified Tobit model seems to be consistent with household's fresh-winter vegetable consumption behavior. This implies that the occurrence of zero expenditures on fresh-winter vegetables results from corner solutions--the household did not desire fresh winter vegetables during the survey period.

REFERENCES

Amemiya, T. "Regression Analysis when the Dependent Variable is Truncated Normal." Econometrica 41(1973): 997-1016.

Amemiya, T., "Tobit Models: A Survey." Journal of Econometrics 24(1984): 3-61.

Amemiya, T., and M. Boskin. "Regression Analysis when the Dependent Variable is Truncated Normal with an Application to the Determinants of the Duration of Welfare Dependence." International Economic Review, 15(1974): 485-96.

Barten, A.P. "Family Composition, Prices and Expenditure Patterns." in P.E. Hart, G. Mills, and J.K. Whitaker (eds.), Econometric Analysis for National Economic Planning. London: Butterworth, 1964.

Becker, G.S. "A Theory of the Allocation of Time." Economic Journal 75(1965): 493-517.

Berndt, E., B. Hall, R. Hall, and J. Hausman. "Estimation and Inference in Nonlinear Structural Models." Annals of Economic and Social Measurement 3(1974):653-665.

Blaylock, J.R., and D.M. Smallwood. "U.S. Demand for Food: Household Expenditures, Demographics, and Projections." U.S. Department of Agriculture, E.R.S., Technical Bulletin No. 1713, February 1986.

Blundell, R., and C. Meghir. "Bivariate Alternatives to the Tobit Model" J. of Econometrics 34(1987): 1979-2000.

Bomberger, W., and D. Denslow. "Some Cross-Sectional Evidence on Demand for Money and other Liquid Assets." Manuscript, University of Florida, 1980.

Burbidge, B.J., L. Magee, and A.L. Rob. " Alternative Transformations to Handle Extreme Values of the Dependent Variable." J. of the American Statistical Association 83(1988): 123-127.

Bureau of Economic and Business Research, Florida Statistical Abstract, Gainesville, Florida: College of Business Administration, University of Florida, 1986.

Buse C.R., and L.E. Salathe "Adult Equivalent Scales: An Alternative Approach." American J. of Agricultural Economics 60(1978): 460-468.

Capps, Jr. O., J.M. Love. "Determinants of Household Expenditure on Fresh Vegetables." Southern J. of Agricultural Economics 15(1983): 127-132.

Chavas, J-P., and A.M., Citzler. "On the Economics of Household Composition." Applied Economics 20(1988): 1401-1418.

Chesher, A. "Testing for Neglected Heterogeneity." Econometrica 52(1984): 865-872.

Cragg, J.G. "Some Statistical Models for Limited Dependent Variables with Applications to the Demand for Durable Goods." Econometrica 39(1971): 829-844.

Deaton, A., and M. Irish. "Statistical Models for Zero Expenditures in Household Budgets." J. of Public Economics 23(1984): 59-80.

Deaton, A., and J. Muellbauer. Economics and Consumer Behavior. New York: Cambridge University Press, 1980.

Engel, E. "Die Lebenskosten Begischer Arbeiter-Familien fruher and jetzt." International Statistical Institute Bulletin 9(1895): 1-74.

Ferber, R. "Consumer Economics, A Survey." J. of Economic Literature 11(1973): 1303-1342.

Fishe, R.P.H., G.S. Maddala, and R.P. Trost. Estimation of a Heteroscedastic Tobit Model. Manuscript, University of Florida, Gainesville, Florida, 1979.

Florida Department of Agriculture, Vegetable Summary 1986-87. Orlando: Agricultural Statics Service, April 1989.

Goldberger, A.S. "Abnormal Selection Bias." in Studies in Econometrics Time Series, and Multivariate Statistics. S. Karlin, T. Amemiya and L. Goodman eds. New York: Academic Press, 1983.

Gorman, W.M. "Tricks With Utility Functions." in M.J. Artis and A.R. Nobay, (ed.) Essays in Economic Analysis: Proceedings of the 1975 AUTE Conference. Sheffield, Cambridge: Cambridge University Press 1976.

Grourieroux, C., A. Monfort, E. Renault and A. Trognon. "Simulated Residuals." J. of Econometrics 34(1987): 201-252.

Greene, W.H., "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model." Econometrica 49(1981):505-513.

Gronau, R., "Wage Comparisons: A Selectivity Bias." Journal of Political Economy 82(1974): 1119-1143.

Haines, P.S., D.K. Guilkey, and B.M. Popkin. "Modeling Food Consumption Decisions as a Two Step Process." American Journal of Agricultural Economics, 70(1988): 543-552.

Heckman, J.J. "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for such Models." Annals of Economic and Social Measurement 5(1976): 475-492.

Huang, C.L., S.M. Fletcher, and R. Raunikar. "Modeling the Effects of the Food Stamp Program on Participating Households' Purchases: An Empirical Application." Southern J. of Agricultural Economics 13(1981): 27-33.

Hurd, M. "Estimation in Truncated Samples when there is Heteroscedasticity." J. of Econometrics 11(1979): 247-58.

Johnson, N.L. "Systems of Frequency Curves Generated by Methods of Translation." Biometrika 36(1949): 149-176.

Kay, J.A., M.J., Keen and C.N. Morris. "Estimating Consumption from Expenditure Data." J. of Public Economics 23(1984): 169-182.

Keen, M.J. "Zero Expenditures and the Estimation of Engel Curves." J. of Applied Econometrics 1(1986): 277-286.

Kokoski, M.F. "An Empirical Analysis of Intertemporal and Demographic Variations in Consumer Preferences." American J. of Agricultural Economics 68(1986): 895-907. S560 .J6 science

Lee, L. and G.S. Maddala. "The Common Structure of Tests for Selectivity Bias, Serial Correlation, Heteroscedasticity and Non-Normality in the Tobit Model." International Economics Review 26(1985): 1-20.

Lewbel, A. "A Unified Approach to Incorporating Demographic or Other Effects into Demand Systems." Review of Economics Studies 52(1985): 1-18.

Lin, T., and P. Schmidt. "A Test of the Tobit Specification Against an Alternative Suggested by Cragg." Review of Economics and Statistics 65(1983): 174-177.

Maddala, G.S. Limited-Dependent and Qualitative Variables in Econometrics. New York: Cambridge University Press, 1983.

Maddala, G.S. A Survey of The Literature on Selectivity Bias as it Pertains to Health Care Markets. Manuscript, University of Florida, Gainesville, Florida, 1985.

Matsumoto, M. Interregional Variation in Food Expenditure Patterns of Low-Income Households. Washington: U.S. Department of Agriculture, Economic Research Service, National Economics Division, June 1984.

Muellbauer, J. "Household Composition, Engel Curves and Welfare Comparisons Between Households." European Economic Review 5(1974): 103-122.

Nelson, F.D. "A Test for Misspecification in the Censored Normal Model." Econometrica 49(1981):1317-1329.

Newey, W. K. "Specification Test for Distributional Assumptions in the Tobit Model." J. of Econometrics 34(1987): 125-145.

Nuckton, F. Carole. Demand Relationships for Vegetables: A Review of Past Studies. Davis: California Agricultural Experiment Station, Giannini Foundation of Agricultural Economics Special Report 80-1, 1981.

Poirier, D.J. "The Use of the Box-Cox Transformation in Limited Dependent Variable Models." J. of American Statistical Association 73(1978): 284-287.

Pollak, R.A., and T.J. Wales. " Estimation of Complete Demand Systems from Household Budget Data: The linear and Quadratic Expenditure Systems." American Economic Review 68(1978): 348-359.

______________."Comparison of the Quadratic Expenditure System and Translog Demand Systems with Alternative Specifications Of Demographic Effects." Econometrica 48(1980): 595-612.

______________."Demographic Variables in Demand Analysis." Econometrica 49(1981): 1533-1551.

330.5 E1911 west

Powell, J.L. "Least Absolute Deviations Estimation for the Censored Regression Model." Journal of Econometrics 25(1984): 303-325.

Powell, J.L. "Symmetrically Trimmed Least Squares Estimation for Tobit Models." Econometrica 54(1986): 1435-1460.

Prais, S.J., and H.S. Houthakker. The Analysis of Family Budgets. 2nd ed., Cambridge: Cambridge University Press, 1971.

Price, D.W., Z.D. Price, and A.D. West. " Traditional and Nontraditional Determinants of Household Expenditures on Selected Fruits and Vegetables." Western J. of Agr. Econ. (1980): 21-35

HD1401 .W47 science

Ramirez, O.A., J.S. Shonkwiler, and R.D. Emerson. Forecasting Variables with Extreme Values: The case of World Sugar Prices. Gainesville: Working Paper, Food and Resource Economics Department, University of Florida, 1988.

Raunikar, R., J.C. Purcell, and J.C. Elrod. Consumption and Expenditure Analysis for Fruits and Vegetables in Atlanta Georgia. Georgia Agricultural Experiment Station, Technical Bulletin NO. 53 June 1966.

Rockwell, G.R., Jr. "Income and Household Size: Their Effects on Food Consumption." U.S. Department of Agriculture, A.M.S. Marketing Research Report No. 340 June 1959.

Salathe, L.E. "The Effects of Changes in Population Characteristics on U.S. Consumption of Selected Foods." American J. of Agricultural Economics 61(1979a): 1036-1045.

Salathe, L.E. Household Expenditure Patterns in the United States. Washington: U.S. Department of Agriculture, Econ. Stat. Coop. Service, Technical Bulletin No. 1603, April 1979b.

Smallwood, D.M., and J.R. Blaylock. "Impact of Household Size and Income on Food Spending Patterns." U.S. Department of Agriculture E.S.S. Technical Bulletin 1650, May 1981.

______________, "Household Expenditures for Fruits, Vegetables, and Potatoes." U.S. Department of Agriculture, ERS Technical Bulletin No. 1690, May 1984.

Smallwood, M.D. and J.M. Harris. "Food Spending in American Households, 1982-84." U.S. Department of Agriculture, E.R.S., Statistical Bulletin No. 753, July 1987.

Sydenstricker, E., and W.I. King. " The Measurement of the Relative Economic Status of Families." Quarterly Publication of the American Statistical Association 17(1921): 842-57.

Tobin, J. "Estimation of Relationships for Limited Dependent Variables." Econometrica 26(1958): 24-36.

U.S. Department of Agriculture, Agricultural Statistics, Washington: U.S. Government Printing Office, 1986.

______________, Economic Research Service. "Food Consumption, Prices, and Expenditures 1963-1983." Statistical Bulletin No. 713. 1985.

United States Department of Commerce, Bureau of the Census. Population Projections of the Population of the United States by Age, Sex and race: 1983 to 2080. Washington: U.S. Government Printing Office, 1984.

______________, Population Estimates and Projections: Projection of the No. of Households and Families: 1986-2000. Washington: U.S. Government Printing Office, 1986.

U.S. Department of Labor, Bureau of Labor Statistics. Consumer Expenditure Survey: (Interview Survey/Diary Survey), Washington: U.S. Government Printing Office, 1980-85.

University of Florida, College of Business Administration, Bureau of Economics and Business Research. Florida Satistical Abstract. Gainesville: The University Press of Florida, 1987.

Varian, H.R. Microeconomic Analysis, 2d ed. New York: Norton, 1984.

White, H. "A Heteroscedasticity-consistent covariance matrix estimator and a direct test for Heteroscedasticity." Econometrica 48(1980): 817-838.

White, H. "Maximum Likelihood Estimation of Misspecified Models." Econometrica 50(1982): 1-25.

BIOGRAPHICAL SKETCH

Anderson Reynolds was born on October 2, 1958, in the town Vieux Fort located at the southern end of the Caribbean island of St. Lucia, West Indies. He graduated from Vieux Fort Secondary School in July 1975.

Immediately following graduation he worked for a year with the St. Lucia Central Water Authority as a laboratory assistant. Subsequently, he enrolled at the Mourne Technical College, Castries, St. Lucia, where he obtained a diploma in general agriculture, November 1977.

After working briefly with WINBAN, a Windward island banana research institute, as an extension officer, he enrolled at Kelsey Institute of Applied Art and Sciences, Saskatoon, Canada, February 1978, where he received a certificate in meat processing, July 1978. Upon his return to St. Lucia he worked for a year with the Ministry of Agriculture as a livestock officer.

His quest for further education, however, led him to Eastern Oklahoma State College in August 1979 to pursue a degree in agricultural economics. After one semester at Eastern he transferred to Southern University, where he spent the following spring and summer semesters. Subsequently, he transferred to Louisiana State University. He received both his Bachelor of Science (May 1982) and Master of Science (August, 1985) degrees from the College of Agriculture, Department of Agricultural Economics, Louisiana State University. During his graduate studies at Louisiana State he served as a graduate research assistant.

He started his Ph.D. studies at the University of Florida, January, 1986. Summer of 1986 he worked with New Jersey's Department of Agriculture, Trenton, New Jersey. He graduated from the University of Florida, August, 1989. While at the University of Florida he worked as a graduate research assistant.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John S. Shonkwiler, Chairman
Professor of Food and Resource Economics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Max R. Langham
Max R. Langham
Professor of Food and Resource Economics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jonq-Ying Lee
Jonq-Ying Lee
Professor of Food and Resource Economics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Leonard K. Cheng
Leonard K. Cheng
Associate Professor of Economics

This dissertation was submitted to the Graduate Faculty of the College of Agriculture and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Jack L. Fry

August, 1989 Dean, College of Agriculture

Dean, Graduate School